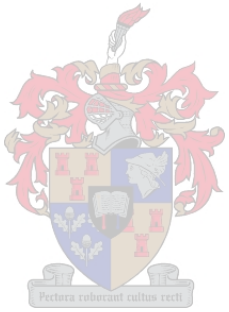
AN INTRODUCTION TO GENERAL RELATIVITY AND ENTROPY BOUNDS

By

Jacques Kotze



Thesis presented in partial fulfilment of the requirements for the degree of Masters of Science at the University of Stellenbosch.

Supervisor : Professor F.G. Scholtz

April 2006

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature

Date

Abstract

Entropy bounds arise from Black hole thermodynamics and are a significant departure from the conventional understanding of the information in a given region. This shift in paradigm is a consequence of the the fact that there is an unexpected relationship between the area and the entropy of a given region of spacetime. Entropy bounds are simplified formulations which are ultimately attempting to be developed into the complete and broad conjecture of the Holographic Principle. This hasn't been achieved successfully as yet. In this thesis the aim is to introduce how the notion of an entropy bound was first suggested and it's subsequent development into more robust formulations. The shortcomings of these conjectures are highlighted along with their strengths.

A foundational introduction of the mathematical requirements for General Relativity is addressed along with an overview of Einstein's theory of gravity. This is illustrated by showing the curvature of relative geodesics as being a consequence of gravity. This is contrasted with Newtonian theory where gravity is also shown to manifests as the curvature of relative geodesics. The working background is concluded with a discussion of Einstein's field equations along with simple and common solutions often used and required.

Opsomming

Swartgat Termodinamika impliseer grense op die entropie, en dus inligting, in 'n gegewe ruimtetyd volume, wat 'n drastiese afwyking van die tradisionele denkwyse oor inligting impliseer. Hierdie paradigma skuif het sy oorsprong in 'n onverwagte verband tussen die oppervlakte van, en entropie bevat, in 'n gegewe ruimte tyd volume. Entropie grense is eenvoudige formulerings van hierdie verwantskap wat uiteindelik beslag moet kry in die volledige en wyer holografiese beginsel. Hierdie doelwit is nog nie bereik nie. Die doel van hierdie tesis is om die oorsprong en verdere formalisering van entropie grense te verduidelik. Beide die sterk en swak punte van die formulerings word bespreek.

Algemene relatiwiteits teorie as 'n teorie van gravitasie, sowel as die wiskundige onderbou daarvan, word oorsigtelik bespreek. Die geometries onderbou van gravitasie word geillustreer aan die hand van die buiging van relatiewe geodete. Dit word met Newton se gravitasie teorie vergelyk wat ook in die buiging van relatiewe geodete gemanifesteer word. Hierdie oorsigtelike agtergrond word afgesluit met 'n oorsig van Einstein se vergelykings, asook eenvoudige en algemene oplossings wat dikwels nodig is en gebruik word.

Acknowledgement

I would like to thank Prof. F.G. Scholtz for his guidance, patience and most importantly for his understanding and compassion when I went through a very trying time. I would further like to thank Prof. L. Susskind for inviting me to Stanford University where I was able to further discuss and get great insight directly from him on this topic. Prof. G.F.R. Ellis was also very accommodating in letting me attend his General Relativity course at U.C.T. To Dr Lisa Dyson I am very grateful for having had the opportunity to learn more about this topic with her and becoming a friend in the process. Similarly to Dr Andrew van Biljon I would like to thank him for his always willing attitude and helpful criticism while always being a good friend. To Lauren Murphy for being loving and supportive throughout I cannot thank you enough for coming into my life and bringing a warmth that I didn't even know I needed. To my friends who brought a balance to my life and always gave me good perspective, I am truly gratefully that you accepted me just as I am and I am a richer person for having you as friends. Lastly to my loving parents and especially my father, who have been through incredible hardship and still always been there for me and encouraged me in all my endeavours. I could not have wished for more caring and special parents.

Contents

List of figures

List of tables

Introduction

General relativity and quantum mechanics are the two foundational pillars in modern physics. Both these disciplines have proved their respective value in the realm of physics as being very secure although there formulation is completely independent of each other. A single theory that integrates these two paradigms still hasn't been developed completely. Investigating and solving the problems that occur when combining these two theories may prove to be a significant aspect in developing a completely unified theory, as discussed by 't Hooft in [tH85]. This is an approach that may lead to uncovering new insight and developing a fresh perspective to better understanding elements of quantum gravity.

One candidate for combining the theories is in the realm of gravitational instabilities. Quantum mechanics has no means for accounting for gravitational effects and thus doesn't have a clear cutoff at which these instabilities occur. Gravitational instability in general relativity, however, are well defined and result in the formation of black holes at sufficiently high energies in a given region of space. Black holes can be evaluated by virtue of their thermodynamic nature from a quantum mechanical perspective, [Bek72], [Bek73], [Bek74], [Haw71] and [Haw72]. Black holes thus pose a natural and interesting setting, since it may elude to some new insights in which to explore the crossover between these two disciplines. The expected result of examining these two theories in this setting is that a cutoff will emerge for gravitational instabilities within quantum mechanics and this will have bearing on quantum gravity since we are dealing with very high energy scales and short length scales (Planck).

The expected result of arriving at a bound is achieved for gravitational instability but the surprising character of this bound is that it is related to the area of the event horizon, [Bek72], [Bek73] and [Bek74]. This means that we are able to relate the number of degrees of freedom, or information, associated with the volume of a black hole to simply an area. This is reminiscent of a Hologram in which all the information of three dimensional object can be captured on a two dimensional surface. This conjecture is thus aptly named the Holographic Principle, [tH93] and [Sus95]. This conjecture is still a work in progress and not well enough established to warrant its place as a fundamental principle in physics. The final chapter of this work looks to explore the reasoning underlying this proposed theory and outlines its development from the generalised second law of thermodynamics, [Bek72], to the suggested covariant entropy bound, [Bou99], which is the closest incarnation of the Holographic Principle to date.

In order to appreciate the arguments in the final chapter an introduction of general relativity is required. The first two chapters accomplish this by covering, in a rudimentary form, an introduction to general relativity. First addressing only essential mathematical requirements to equip the reader the necessary tools for tackling Einstein theory of gravity. This is introduced in chapter two and focuses on curvature being a means for understanding gravitational effects. This is contrasted with our understanding of Newtonian gravity and reformulates it so that it too is cast in a form where curvature is used to describe the attractive forces between matter. This then culminates in the stating of Einstein's field equation and some corresponding common solutions are shown.

Chapter 1
Spacetime Basics

General relativity is an enhanced extension of special relativity since it caters for gravity.This is done successfully in general relativity by abandoning the naive perception of space-time being flat and considers relativity on a curved manifold. In general relativity the physics of a freely falling observer carrying locally flat coordinates is locally indistinguishable from that of special relativity. This is Einstein's famous strong equivalence principle. This allows us the framework to understand how physics works in an inertial frame while also making it possible to locally implement the tools already crafted in special relativity.

The mathematical structure of this 4-dimensional non-Euclidean geometry is referred to as a pseudo-Riemannian manifold. In order to generate a successful understanding of general relativity we must address the mathematical demands of working on a geometry of this nature, which falls in the domain of Differential Geometry. Differential geometry gives a comprehensive structure within which to approach the necessities of general relativity. In particular the principle of invariance must be kept intact by formulating the laws in a way that is valid independent of the coordinate system used to label spacetime. This is the essence of relativity and is embodied in the foundation of differential geometry as tensors. Tensors were conceived in the setting of differential geometry but find themselves very suitable adept to the requirements of general relativity.

Tensors, in their broadest sense, encompass all operators that are linear in nature. In order to generate a natural understanding of tensors most literature builds on the conceptual basis intrinsic to scalars and vectors. The fundamental understanding of tensors is innately known by any student who has come across scalars and vectors since these are tensors of the most simple type. More complex or higher rank tensors are merely a natural progression of these primary concepts of scalars and vectors.

The aim of this chapter is to introduce the foundation of general relativity by introducing the most essential mathematics required to gain a working knowledge of general relativity. This is accomplished by taking a hands on and bare bones approach in which the reader is guided through a logical set of steps beginning with coordinates and then progressively building on top of this to develop a comprehensive means by which to view the world. In the following chapter more complex tools will be crafted to manipulate and interpret this world view and thus examine the nature of gravity.

Beginning with an introductory review of coordinates, events, curves and surfaces, this is later combined and their interplay examined, culminating in an illustration of standard type calculations that are commonly required given these fundamental building blocks. The important concept of coordinate transformation is then introduced and applied to the events, curves and surfaces discussed earlier. This is followed by a section focusing on tensors and their transformation laws contrasted with what has already been introduced for coordinates and a penultimate discussion focuses on tensor operations and how in general tensors can be manipulated. Finally bases are reviewed during which a general basis, coordinate basis and tetrad basis are covered.

The structure of this chapter follows the outline of the general relativity course taught by Prof. Ellis, [Ell99].

This chapter serves as a grounding for more complex topics introduced later on and in so doing has a very hands on approach, always trying to progressively develop further on concepts that have been introduced. This idea is continued into the second chapter where gravity is introduced by means of measuring its affects on free falling observers. For this purpose the concept of a covariant derivative and a geodesic is required and presented at the beginning of this chapter.

1.1 Notational Convention

Throughout this discussion Planck units are used:

$$\hbar = G = c = k = 1, \tag{1.1}$$

where G is Newton's constant, $\hbar$ is Planck's constant, c is the speed of light, and k is Boltzmann's constant. In particular all areas are measured in multiples of the square Planck length,

$$l_{\mathrm{P}}^2 = \frac{G\hbar}{c^3} = 2.59 \times 10^{-66}\mathrm{cm}^2. \tag{1.2}$$

The notational convention followed where N is the number of degrees of freedom and $\mathcal{N}$ is the dimensionality of the Hilbert space.

1.2 Spacetime

Spacetime is composed of three primary components: Events, curves and surfaces. Each of these components are important on their own, as is the interplay between them. The following section aims to give a hands on description of these components and builds on each element one at a time and then aims to evaluate their common relationship.

1.2.1 Coordinates

Spacetime represent the "arena" of physics. This is a set of events, places and times, where objects move and interact. This arena is a four dimensional manifold, M, which is locally the same as a four dimensional Cartesian space. Coordinates represent a logical and coherent way in which to record or document the events in the arena mentioned above and are generally composed of sets of four, for example (t, x, y, z) or (t, r, θ, ϕ).

Einstein's summation convention is used in the notation $x^i = (x^0, x^1, x^2, x^3)$ since it is more economical in describing complex equations and also a standard form convention in literature on the topic of general relativity. Two points that should be kept in mind when considering coordinates is that one can use any coordinates which best suite the situation and one cannot, in general, cover the whole of spacetime by one coordinate system. Overlapping coordinates must often be used and only in very special cases is one coordinate system sufficient.

In the notation described above a four vector would be described as $X^i = (X^0, X^1, X^2, X^3)$

while a tensor of rank two in $1+1$ dimensions can be represented as a 2×2 matrix,

$$T_{ab} = \begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix}. \tag{1.3}$$

Some points to be kept in mind in the following sections:

- In general an n-dimensional space is considered, however examples are kept to two and three dimensions for the purpose of simplification and illustration.
- The Einstein summation convention is always assumed.
- The Kronecker delta δ^i_j is defined as

$$\delta^i_j = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}$$

The fundamental elements composing spacetime are events, curves and surfaces.

1.2.2 Events

A loose analogy of what an event encompasses is a point in a snap shot or photo, taken to capture space at an instance in time. This is achieved by labelling each of the spacetime points with a chosen set of coordinates. This is a local map ϕ from the manifold M to the coordinate space $\mathbb{R}^4$.

This allows us to assign to each event P the corresponding coordinate $x^i(P)$. Since this is a 1-1 mapping we can refer to P uniquely. This also helps us determine what are open and closed sets in M and thus fixes the spacetime topology, continuity and differentiability properties of the spacetime.

To summarise, a spacetime is simply a set of all the physically connected events and we may need several coordinate charts in order to cover the spacetime successfully.

1.2.3 Curves

A curve in spacetime is a one dimensional set of events denoted by $x^i(v)$ where v is the curve parameter. To follow the analogy from the previous section, if an event is a photo, a curve would be a moving picture or movie, which would portray a sequence of events or a story. This represents a local map from $\mathbb{R}^1$ into M. An example of a curve would include the history of a particle or otherwise referred to as the particle worldline. Each curve is associated with a tangent vector which gives us its direction in spacetime at each point on the curve. The components of this vector are

$$X^i = \frac{dx^i}{dv}. \tag{1.4}$$

If we are considering particle motion the above equation would denote a representative four velocity.

A family of curves that fills up spacetime with tangent vectors defined at each point, is called a congruence of curves and the tangent vector gives the fluid four velocity at each point. We can determine the family of curves through the use of equation (1.4).

1.2.4 Surfaces

A single surface is a three dimensional set of events. It can be described by a non-constant function $f(x^i)$ through the equation $f(x^i) = c$ where c is a constant. A surface can thus be seen as a cross section or slice of spacetime. A family of surfaces can be obtained by varying c in an interval I of R. This family represents a non-intersecting group that is a map from M to $\mathbb{R}^1$. There is a set of 4 coordinates x^i and thus a corresponding family of four independent surfaces in spacetime.

1.2.5 Curves and surfaces relation

We can establish a relationship between curves and surfaces by taking into account the equation for a curve $x^i = x^i(v)$ and applying it to the equation for a family of surfaces $f = f(x^i)$ resulting in $f(v) = f(x^i(v))$. The rate of change of this relation is given by

$$\frac{df}{dv} = \frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial v} = X^i\frac{\partial f}{\partial x^i} \tag{1.5}$$

where

$$\mathbf{X} = X^i(x^j)\frac{\partial}{\partial x^i} \tag{1.6}$$

is a directional derivative operator in the direction X.

When we consider the vector $\mathbf{X}$ from a geometric perspective it is considered an arrow, analytically however, it is considered to be a differential operator and is expressed in terms of the basis directional derivatives $\frac{\partial}{\partial x^i}$ with components of the form $X^i(x^j)$. Thus the analytic viewpoint sees the operators $\mathbf{X}$ as being directional derivatives which uniquely characterise a direction at each point.

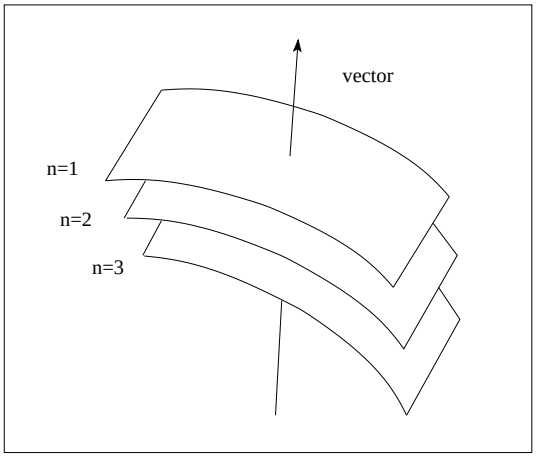


Figure 1.1: Illustration of a tangent vector, X, passing through a function of planes, $f = c + \Delta n$. The vector field, $X(f)$, is a function that counts the number of planes the vector transvects.

The vector field $\mathbf{X}$ can be seen to map functions f to $X(f)$ from the following

$$\frac{df}{dv} = X^i \frac{\partial f}{\partial x^i} =: X(f).$$

We can understand this mapping in geometric terms by associating the function f to a set of planes or level surfaces, $f = c + \Delta n$, which are being crossed by the curve with tangent vector X. Therefore $X(f)$ would be larger if more surfaces were being transvectored. Keeping count of the number of surfaces being crossed by the curve would be a means of determining the rate at which the function f changes. This overall process would be equivalent to the mapping from f to $X(f)$. Figure 1.1 shows how the description above can be visualised.

The variation of the function is characterised analytically by differential df with components

$$df_i = \frac{\partial f}{\partial x^i} \iff X(f) = df_i X^i =: df(\mathbf{X}) \tag{1.7}$$

for any vector field $\mathbf{X}$

$$\mathbf{d}f : \mathbf{X} \to df(\mathbf{X}) \equiv X(f) \tag{1.8}$$

which can be seen as two nearby planes or level surfaces.

1.2.6 Coordinates, curves and surfaces

A $1-1$ coordinate map exists, $\phi : M \to \mathbb{R}^n$ which maps the surfaces and curves in $\mathbb{R}^n$ into the space M. This is relevant in understanding the coordinates which are affected by this mapping by virtue of their association with a family of surfaces, curves and vectors. We can thus have a better appreciation for the coordinates by examining the related curves and surfaces.

The coordinate surfaces are defined by $\{x^i = \text{const}\}$ for $i = 1$ to n. The coordinate curves lie in the intersection of these surfaces, for example the curve $x^i(v)$ given by $x^1 = v, x^2 = \text{const}, x^3 = \text{const}$ in $\mathbb{R}^3$ lies in the intersection of the surface $x^2 = \text{const}, x^3 = \text{const}$. By Eq. (1.4), its tangent vector has components $X^i = (1, 0, 0) = \delta^i_1$, this implies that x^1 is the curve parameter and that x^2 and x^3 do not vary along the curve. From Eq. (1.6) its tangent vector is $\mathbf{X} = \delta^i_1 \frac{\partial}{\partial x^i} = \frac{\partial}{\partial x^1}$. Similarly the curve $\{\text{only } x^i \text{ varies}\}$ implies x^j ($\text{where} j \neq i$) is constant, and this is the curve tangent vector $\mathbf{X} = \frac{\partial}{\partial x^i} \iff X^j = \delta^j_i$ where i is fixed and j runs from 1 to n. Thus associated with the coordinate system there is a set of n vector fields $e_k = \frac{\partial}{\partial x^k}$, with components $e^i_k = \delta^i_k$, where k labels the vector field and i labels the component. This is the coordinate basis associated with the coordinates $\{x^i\}$. Eq. (1.6) expresses the vector field $\mathbf{X}$ in terms of this basis.

From Eq. (1.5) using $f = x^j$, we obtain

$$\frac{dx^j}{dv} = X(x^j) = X^i \frac{\partial x^j}{\partial x^i} = X^i \delta^j_i = X^j, \tag{1.9}$$

which simply tells us that the rate of change of the coordinates x^j along the curve with tangent vector X is equal to the components X^j of the vector field X.

There exists a similar relationship between the coordinates and there associated differentials.

Using Eq. (1.7) where $f = x^j$ results in

$$d(x^j)_i = \frac{\partial x^j}{\partial x^i} = \delta^j_i \iff X(x^j) = dx^j(\mathbf{X}) = X^j$$

where the i components of the first differential is zero unless $i = j$, while in the right hand equation the differential dx^j is just the mapping of the vector field $\mathbf{X}$. These differentials, $d(x^j)$, form a coordinate basis of differentials. They are the dual basis to the basis of the vectors, $\frac{\partial}{\partial x^i}$. This can be shown by applying Eq. (1.7) to the coordinate basis of vectors; we find $\frac{\partial}{\partial x^j}(x^i) =: dx^j(\frac{\partial}{\partial x^i}) = \delta^j_i$. Geometrically this corresponds to the fact that the coordinate curves and the surfaces exactly fit together.

Using Eq. (1.7) and given an arbitrary function, $f(x^j)$, produces

$$\mathbf{d}f = df_i \mathbf{d}x^i \iff df(\mathbf{X}) = df_i dx^i(\mathbf{X}) = \frac{\partial f}{\partial x^i} X^i \tag{1.10}$$

where the first equality express the differential $\mathbf{d}f$ in terms of the basis differential $\mathbf{d}x^i$, with components $df_i(x^k)$. The right hand equality explicitly shows what this map is when applied to the arbitrary vector field $\mathbf{X}$. If we choose $\mathbf{x} = \frac{\partial}{\partial x^k}$ for some coordinate x^k, we find the k'th component of the differential, $df_k = df(\frac{\partial}{\partial x^k})$.

1.2.7 Standard Calculations

The standard type of calculations we may require to perform can range from determining angles to distances. In a three dimensional Euclidean space we would normally consider a distance being represented by a small coordinate displacement $dx^i = (dx, dy, dz)$ where

$$ds^2 = dx^2 + dy^2 + dz^2,$$

this implies that when considering a curve $x^i(v)$ the distance would be calculated by using

$$S = \int_P^Q \sqrt{ds^2} = \int_P^Q \sqrt{dx^2 + dy^2 + dz^2} = \int_P^Q \sqrt{\left(\frac{dx}{dv}\right)^2 + \left(\frac{dy}{dv}\right)^2 + \left(\frac{dz}{dv}\right)^2}\, dv.$$

If we choose a more general coordinate system we require the metric tensor to calculate the displacement component

$$ds^2 = g_{ij}(x^k) dx^i dx^j \tag{1.11}$$

where the metric is symmetric. This can then be used to produce the familiar scalar product of two vectors

$$X \cdot \mathbf{Y} = g_{ij} X^i Y^j \Rightarrow X^2 = g_{ij} X^i X^j = |\mathbf{X}|^2, \tag{1.12}$$

with the respective angles between the vectors being calculated with

$$\cos\theta = \frac{\mathbf{X} \cdot \mathbf{Y}}{|\mathbf{X}||\mathbf{Y}|} \tag{1.13}$$

with a zero result implying that the vectors $\mathbf{X}$ and $\mathbf{Y}$ are orthogonal. In a more general setting where we consider a curved spacetime Eq. (1.11) can be used to determine

$$S = \int_P^Q \sqrt{ds^2} = \int_P^Q \sqrt{g_{ij}(x^k)dx^i dx^j} = \int_P^Q \sqrt{g_{ij}\frac{dx^i}{dv}\frac{dx^j}{dv}}\, dv \tag{1.14}$$

by using the metric and the scalar product.

There may occur coordinate singularities during these calculations. These are not physical singularities and are as a result of poor coordinate choices. Spherical coordinates exhibit this singularity at $r = 0$, for example. These are local phenomena and this is why several coordinate patches are usually required.

We can implement the above for a flat spacetime and recover the familiar relations associated with special relativity. The metric tensor is $\eta_{ij} = \text{Diag}(-1, 1/c, 1/c, 1/c)$, resulting in

$$ds^2 = -dt^2 + (dx^2 + dy^2 + dz^2)\frac{1}{c^2}, \tag{1.15}$$

where c is the speed of light and is often set to 1 with the appropriate choice of units. The above equation can be re-written in the form

$$ds^2 = dt^2\left(-1 + \frac{1}{c^2}\left(\frac{dr}{dt}\right)^2\right) = -dt^2\left(1 - \frac{v^2}{c^2}\right) \tag{1.16}$$

where $dr^2 = dx^2 + dy^2 + dz^2$ and the local speed of motion is v. More information can be extracted from the spacetime interval by using the following properties:

Properties:

- $ds^2 < 0$ which implies that $v^2 < c^2$: This represents a spacetime path with a proper time τ along the particle world line

$$\begin{aligned} \tau &= \int_P^Q \sqrt{|ds^2|} = \int_P^Q \sqrt{-g_{ij}(x^i)dx^i dx^j} \\ \therefore d\tau &= \sqrt{-ds^2}. \end{aligned} \tag{1.17}$$

  Using Eq. (1.16) we find

$$\begin{aligned} \frac{d\tau}{dt} &= \sqrt{1 - \frac{v^2}{c^2}} \\ \therefore \gamma(v) = \frac{dt}{d\tau} &= \frac{1}{\sqrt{1 - v^2/c^2}}, \end{aligned} \tag{1.18}$$

  which is in accordance with the well known relation of Special Relativity.

- $ds^2 = 0$: This implies the motion is at the speed of light i.e. $v^2 = c^2$.

- $ds^2 > 0$ which implies $v^2 > c^2$: We calculate the proper distance D along a path by

$$D = \int_P^Q \sqrt{|ds^2|}$$
$$\therefore dD = \sqrt{ds^2}. \tag{1.19}$$

- The metric also determines simultaneity for an observer which corresponds to spacetime orthogonality. If we consider a metric of the same form as the flat spacetime metric, $\eta_{ij} = \text{Diag}(-1, 1, 1, 1)$ and a tangent vector T^i to the timelike curve (where only t varies) it would have components $T^i = \delta^i_0$, while any curve tangent to the surfaces of simultaneity (where $t = \text{const}$) for an observer moving on these world-lines, has tangent vector $Y^i = (0, Y^\nu)$ which in turn implies that $Y(t) = 0$. We therefore have $\mathbf{T} \cdot \mathbf{Y} = T^i g_{ij} Y^j = 0$. This determines that the displacement Y^i is instantaneous for an observer with four velocity T^i.

1.3 Changing coordinates

We want to be able to change to any arbitrary set of coordinates and still retain the validity of the equations. First we must establish how we will represent a change in the coordinates. This is done by choosing a mapping from an old set of coordinates x^i to a new set of coordinates $x^{i'}$. Expressing the new coordinates of each point in terms of the old, is called a forward transformation and is given by

$$x^{i'} = x^{i'}(x^i). \tag{1.20}$$

The inverse relation is also applicable and is called the backward transformation

$$x^i = x^i(x^{i'}) \tag{1.21}$$

giving the old coordinates at each point in terms of the new coordinates.

1.3.1 Functions

The transformation of coordinates, understandably, changes all the representations of objects in M. This then has a knock on effect for the function $f(x^i) \Rightarrow f(x^{i'}) = f(x^{i'}(x^i))$. This change in the coordinates results in the form of the function changing and should strictly have a different notation, i.e. $\tilde{f}$, but instead we choose to keep it the same and only show the change in the coordinates transformation.

1.3.2 Curves and vectors

If we consider a curve, denoted by $x^i(\lambda)$ then the resulting curve under the transformation discussed above is

$$x^{i'} = x^{i'}(x^i(\lambda)),$$

with the tangent vectors new components given by

$$X^{i'} = \frac{d}{d\lambda}(x^{i'}) = \frac{d}{d\lambda}\left(x^{i'}(x^i(\lambda))\right) = \frac{\partial x^{i'}}{\partial x^i}\frac{dx^i}{d\lambda} = \frac{\partial x^{i'}}{\partial x^i}X^i,$$

having used the chain rule. This means that the forward transformation is

$$X^{i'} = A^{i'}{}_{i} X^{i}, \tag{1.22}$$

where

$$A^{i'}{}_{i} = \frac{\partial x^{i'}}{\partial x^{i}}. \tag{1.23}$$

The inverse of this transformation can be used to establish the backward transformation of the tangent vector

$$X^{i} = A_{i'}{}^{i} X^{i'}, \tag{1.24}$$

where

$$A_{i'}{}^{i} = \frac{\partial x^{i}}{\partial x^{i'}}. \tag{1.25}$$

This then demands a relationship between the transformations as being defined as, $A_{i'}{}^{i} A^{i'}{}_{j} = \delta^{i}_{j}$ and $A^{i'}{}_{i} A_{j'}{}^{j} = \delta^{i}_{j}$. An alternative way in which to obtain the vector transformation law is to note that we can write the vector $\mathbf{X}$ equally well in terms of the old and new basis by

$$\mathbf{X} = X^{i} \frac{\partial}{\partial x^{i}} = X^{i'} \frac{\partial}{\partial x^{i'}}. \tag{1.26}$$

By using the partial differential chain rule,

$$\frac{\partial}{\partial x^{i'}} = \frac{\partial x^{i}}{\partial x^{i'}} \frac{\partial}{\partial x^{i}},$$

we can re-write Eq. (1.26) as

$$\mathbf{X} = X^{i} \frac{\partial}{\partial x^{i}} = X^{i'} \frac{\partial x^{i}}{\partial x^{i'}} \frac{\partial}{\partial x^{i}}.$$

This, upon comparison, leads to the same transformation law as in Eq. (1.24).

Let us illustrate the above with a 2-dimensional example where $x^{i}(x^{i'})$ takes the form $x = r\cos\theta$ and $y = r\sin\theta$.

The transformation matrix $A_{i'}{}^{i}$ has the form

$$A_{i'}{}^{i} = \frac{\partial x^{i}}{\partial x^{i'}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix}.$$

Therefore the tangential vector is

$$X = \frac{\partial}{\partial \theta} \iff X^{i'} = \delta^{i'}_{2} = (0, 1)$$

and its transformation is

$$X^{i} = X^{i'} A_{i'}{}^{i} = \delta^{i'}_{2} A_{2'}{}^{i} = \frac{\partial x^{i'}}{\partial x^{2'}},$$

giving the result

$$X^{i} = (-r\sin\theta, r\cos\theta).$$

We can now find the inverse transformation matrix by inverting the Jacobian matrix obtained

above

$$A^{i'}{}_{i} = \begin{pmatrix} \cos\theta & -\sin\theta/r \\ \sin\theta & \cos\theta/r \end{pmatrix}.$$

This gives the inverse transformation matrix in terms of the new coordinates and in order to write it in terms of the old coordinates, we must substitute for the original relation, $x = r\cos\theta$ and $y = \sin\theta$, to obtain

$$A^{i'}{}_{i} = \begin{pmatrix} x/(x^2+y^2)^{1/2} & -y/(x^2+y^2)^2 \\ y/(x^2+y^2)^{1/2} & x/(x^2+y^2) \end{pmatrix}.$$

1.3.3 Gradient of function

As we saw with the change in coordinates leading to a change in the form of the function, so also it will lead to a change in the differential form,

$$df_{i'} = \frac{\partial f}{\partial x^{i'}} = \frac{\partial x^i}{\partial x^{i'}}\frac{\partial f}{\partial x^i} = \frac{\partial x^i}{\partial x^{i'}} df_i.$$

We can then single out the transformation from the above as follows

$$df_{i'} = A_{i'}{}^{i} df_i, \quad df_i = A^{i'}{}_{i} df_{i'}. \tag{1.27}$$

1.3.4 Invariants

We require a method by which we can make certain functions invariant, or put another way, let functions take the same value for all coordinate systems. A way of achieving this is to use the metric to construct a scalar product which maps the pair of vectors $(x_\mu, x_v) \to g^{\mu v} x_\mu x_v \in \mathbb{R}$. If the metric transforms appropriately invariance can be achieved. Writing the above requirements that the scalar product remain invariant regardless of the coordinate system, results in the following relation

$$ds^2 = ds'^2 \iff g_{ij} dx^i dx^j = g_{i'j'} dx^{i'} dx^{j'}, \tag{1.28}$$

which leads us to establishing the transformation conditions for the metric,

$$\begin{aligned} g_{ij} dx^i dx^j &= g_{i'j'} A^{i'}{}_{i} dx^i A^{j'}{}_{j} dx^j \\ \Rightarrow Z_{ij} dx^i dx^j &= 0, \quad \text{where} \quad Z_{ij} := g_{ij} - g_{i'j'} A^{i'}{}_{j} A^{j'}{}_{j}. \end{aligned} \tag{1.29}$$

The independence of dx^i and dx^j implies that $Z_{ij} = 0$, i.e. ,

$$g_{ij} = g_{i'j'} A^{i'}{}_{i} A^{j'}{}_{j} \quad \iff \quad g_{i'j'} = g_{ij} A_{i'}{}^{i} A_{j'}{}^{j}, \tag{1.30}$$

which leaves us with the derived transformation law of the metric tensor components.

1.4 Tensors and Tensor transformations

A tensor transformation is merely a generalisation of the previous transformations we have discussed for vectors and scalars, which are tensors of rank zero and one, respectively.We would thus expect that the transformations for tensors would be very similar to those for vectors.

An example would be if we used a tensor of the form $R^{ijk}{}_{gh}$ which would transform as

$$R^{i'j'k'}{}_{g'h'} = R^{ijk}{}_{gh} A^{i'}{}_{i} A^{j'}{}_{j} A^{k'}{}_{k} A_{g'}{}^{g} A_{h'}{}^{h}. \tag{1.31}$$

In terms of a mechanical way of viewing the previous equation we can see that the transformation matrix acts as a substitution operator. In every case it cancels the old index, by summation, and replaces it with a new index in the same place.

Since the coordinates transformations satisfy the properties of a group, these properties then also extend to the transformation matrices. They can be categorised as

- The identity transformation $x^i \to x^{i'} = x^i$ leads to the unit matrix $A^{i'}{}_{i} = \delta^{i'}{}_{i}$, giving the identity tensor transformation in Eq. (1.31).
- The inverse transformation matrix $A_{i'}{}^{i}$ is used to make an inverse or backward transformation as illustrated in Eq. (1.21).
- Two transformations can be composed together to form a single new transformation i.e. $T_1 : x^i \to x^{i'}$ where $x^{i'} = x^{i'}(x^i)$ and $T_2 : x^{i'} \to x^{i''}$ where $x^{i''} = x^{i''}(x^{i'})$, leads to the overall transformation $T_3 : x^i \to x^{i''}$ where $x^{i''}(x^i) = x^{i''}(x^{i'}(x^i))$, with the corresponding composite tensor transformation given by the composition formula $A^{i''}{}_{i} = A^{i''}{}_{i'} A^{i'}{}_{i}$, having used Eq. (1.23). This is equivalent to implementing the chain rule for partial derivatives.

Thus the set of tensor transformations form a local group with the above properties.

1.5 Tensor equations

The most important attribute about tensors is that if a tensor equation is true in one coordinate system, it is true in all coordinate systems. In the following example

$$T_{ij} = R_{ij} \tag{1.32}$$

if both are tensors in an initial coordinate system, where T_{ij} and R_{ij} are components of tensors, then

$$T_{i'j'} = A_{i'}{}^{i} A_{j'}{}^{j} T_{ij} = A_{i'}{}^{i} A_{j'}{}^{j} R_{ij} = R_{i'j'}. \tag{1.33}$$

Thus $T_{i'j'} = R_{i'j'}$ in every coordinate system.

This is a characteristic that is in particular sought after in general relativity, where we require equations of physical significance to remain unchanged regardless of the coordinate system chosen. This is a loose definition of covariance.

1.5.1 Tensor operations

There are a couple of key operations that we need to become familiar with in order to be able to manipulate tensors and create new tensors.

Linear combination:

A linear combination of two or more tensors is a new tensor. An example would be

$$S^{ij}{}_{c} = \alpha T^{ij}{}_{c} + \beta R^{ij}{}_{c},$$

T^{ij} and R^{ij} are existing tensors along with α and β being constants, either positive or negative, all combining to form a new tensor S^{ij}.

Multiplication:

Similarly we can multiply tensors together to produce a new tensor, i.e.

$$R^{ij}{}_{c}S^{ef}{}_{gh} = K^{ij\ ef}{}_{c\ gh}.$$

This operations then simply amounts to keeping track of indices of the respective tensors being multiplied together. It is worth noting that multiplying functions with tensors also produce new tensors. Thus $T^{ij} = fR^{ij}$ produces a new tensor from a function and tensor.

Contraction:

In order to contract a tensor you require at least one upper index and another lower index, these then cancel each other out over the summation. For example contracting $R^{ij}{}_{cd}$ on indices i and c,

$$R^{ij}{}_{cd} \rightarrow T^{j}{}_{d} = R^{ij}{}_{id}. \tag{1.34}$$

When a contraction takes place over two indices, it means that the summation of m indices reduces to $m-2$ indices. Of course if we contract all the indices of the tensor we produce a scalar. A scalar produced in this way is coordinate independent and thus invariant, which is useful in certain circumstances.

$$T^{i}{}_{j} \Rightarrow T := T^{i}{}_{i} \tag{1.35}$$

Raising and Lowering indices:

One of the simplest and most common operations required to perform on tensors is a lowering or raising of indices. This is a transformation from the vector space to the dual space or visa versa. The metric tensor is used to contract the tensor requiring the raising or lowering. For example

$$X_i = g_{ij}X^j. \tag{1.36}$$

Similarly we use the inverse metric tensor, g^{ij}, with indices upstairs. We can do this since the metric tensor is non-degenerate i.e.

$$g^{ij}g_{jk} = \delta^{i}_{k}. \tag{1.37}$$

We thus use the inverse metric to raise indices in much the same way as we lower the indices earlier

$$X^i = g^{ij}X_j. \tag{1.38}$$

Using the metric tensor and the inverse metric tensor makes it possible for us to write the indices in any configuration. By facilitating this operation it also then allows us to easily produce scalar products.

$$\mathbf{A}\cdot\mathbf{B} = A^i g_{ij} B^j = A^i B_i = A_j B^j \tag{1.39}$$

In this instance of forming scalars, the position of the indices being up or down becomes immaterial.

Let us look at the special instance of raising and lowering indices on the metric tensor. If we

raise one index on the metric tensor g_{ij}, we obtain $g^i{}_j = g^{ik}g_{kj} = \delta^i_j$ and in a similar way we can raise the second index, $g^{ik}g_{km}g^{jm} = g^{ij}$. We can thus conclude that $g_{ij}, g^i{}_j, g^{ij}$ are in all aspects the same tensor for the spacetime. An interesting point that we can highlight is that the scalar $g^a{}_a$ is equal to the dimension of the space e.g. $g^a{}_a = \delta^a_a = \delta^0_0 + \delta^1_1 + \delta^2_2 + \delta^3_3 = 1 + 1 + 1 + 1 = 4$.

1.5.2 Symmetry properties

Symmetric properties are very important and can be used to simplify tensor equations properties. Transformations of tensors preserve symmetries.

A tensor can be written as the sum of its symmetric and antisymmetric parts,

$$T_{ij} = \frac{1}{2}(T_{ij} + T_{ji}) + \frac{1}{2}(T_{ij} - T_{ji}) = T_{(ij)} + T_{[ij]}, \tag{1.40}$$

where $T_{(ij)} = \frac{1}{2}(T_{ij} + T_{ji})$ is the symmetric part and $T_{[ij]} = \frac{1}{2}(T_{ij} - T_{ji})$ is the skew symmetric part. The above tensor is therefore symmetric if $T_{ij} = T_{(ij)}$ which implies that $T_{ij} = T_{ji}$. It is however skew-symmetric if $T_{ij} = T_{[ij]}$ which implies that $T_{ij} = -T_{ji}$.

1.6 General Bases

It is worthwhile considering a basis in which the magnitude of the displacement is given for each of the components. This is called a physical basis. For this one requires that the corresponding metric components are unity and that the basis vectors are orthogonal to each other. This is best illustrated in an example where we take a coordinate basis and convert it to an orthonormal basis.

For an Euclidean plane in polar coordinates, $ds^2 = dr^2 + r^2 d\theta^2$ where $g_{ij} = \text{Diag}(1, r^2)$. This in turn produces the magnitude of the coordinate basis vectors, $g_{rr} = (\frac{\partial}{\partial r} \cdot \frac{\partial}{\partial r}) = 1$ and $g_{\theta\theta} = (\frac{\partial}{\partial \theta} \cdot \frac{\partial}{\partial \theta}) = r^2$ with scalar product $g_{r\theta} = (\frac{\partial}{\partial r} \cdot \frac{\partial}{\partial \theta}) = 0$. We now choose an orthonormal basis e_a by dividing each vector by its magnitude,

$$\mathbf{e}_r = \frac{\partial}{\partial r}, \quad \mathbf{e}_\theta = \frac{1}{r}\frac{\partial}{\partial \theta}, \tag{1.41}$$

then we have achieved a unit vector representation where $\mathbf{e}_r \cdot \mathbf{e}_r = (\frac{\partial}{\partial r} \cdot \frac{\partial}{\partial r}) = g_{rr} = 1$, $\mathbf{e}_\theta \cdot \mathbf{e}_\theta = g_{\theta\theta} = (\frac{1}{r})^2(\frac{\partial}{\partial \theta} \cdot \frac{\partial}{\partial \theta}) = (\frac{1}{r})^2 g_{rr} = \frac{1}{r^2}r^2 = 1$ and also showing that they are orthogonal $\mathbf{e}_r \cdot \mathbf{e}_\theta = \frac{1}{r}\frac{\partial}{\partial r} \cdot \frac{\partial}{\partial \theta} = \frac{1}{r}g_{r\theta} = 0$. This then adheres to the requirements of an orthonormal basis $g_{ab} = \text{Diag}(1, 1)$.

The transformations above take the form

$$\mathbf{e}_a = \Lambda_a{}^i(x^k)\frac{\partial}{\partial x^i} \iff \frac{\partial}{\partial x^i} = \Lambda^a{}_i(x^k)e_a \tag{1.42}$$

where

$$\Lambda_a{}^i = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r} \end{pmatrix}$$

and the inverse is

$$\Lambda^a{}_i = \begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix}.$$

This is very reminiscent of the equation following Eq. (1.26) which also indicates that the components of vectors will transform in a natural way from the above

$$\mathbf{X} = X^a \mathbf{e}_a = X^i \frac{\partial}{\partial x^i} \tag{1.43}$$

which shows that

$$X = X^a \Lambda_a{}^i(x^k) \frac{\partial}{\partial x^i} = X^i \frac{\partial}{\partial x^i}$$

and so

$$X^a = \Lambda^a{}_i X^i \iff X^i = \Lambda_a{}^i X^a. \tag{1.44}$$

As with the transformations we have already applied to tensors in Eq. (1.31) and its inverse, these equations still hold in the same way for vectors in general, but just with the new transformation matrix $\Lambda_a{}^i$ instead of $A_{i'}{}^i$.

The above example illustrates that for an arbitrary displacement we have a simplified metric form, where

$$\eta^a = \Lambda^a{}_i(x^k) dx^i \iff dx^i = \Lambda_a{}^i(x^k) \eta^a \tag{1.45}$$

is given as

$$ds^2 = g_{ab} \eta^a \eta^b = g_{ab} \Lambda^a{}_i(x^k) dx^i \Lambda^b{}_j(x^k) dx^j, \tag{1.46}$$

and since we are dealing with an orthonormal basis this becomes

$$ds^2 = (\eta^1)^2 + (\eta^2)^2.$$

The downside to this simplification is that we may no longer be dealing with a coordinate basis which means that there doesn't exist any coordinates z^1 and z^2 such that $e_a = \frac{\partial}{\partial z^a}$. In a similar vein it is not possible to express a displacement $\eta^a = \Lambda^a{}_i(x^k) dx^i$ as an exact differential $\eta^a = dz^a$ and subsequently we cannot express the metric as $ds^2 = (dz^1)^2 + (dz^2)^2$. From Eq. (1.45) and Eq. (1.46) we have the following form of the metric

$$ds^2 = (\Lambda^1{}_i(x^k) dx^i)^2 + (\Lambda^2{}_j(x^k) dx^j)^2$$

which expresses the local Pythagorean distance relation for the space.

Thus vectors transform just as before but with the transformation matrix $\Lambda_a{}^i$ instead of $A_i{}^i$. The previous discussion about tensor transformation holds as before but just with the new transformation matrix in Eq. (1.31) and its inverse.

This example shows the possibility of using a far more general basis than coordinate bases. They have advantages which are often exploited and explain their common use in General Relativity. The next sections explore changing to a completely general bases and then consider the properties of coordinate and tetrad bases.

1.6.1 General Basis

Changing a coordinate basis to a general basis $\{\mathbf{e}_a\}$ can be achieved by using Eq. (1.42). We can furthermore change from one general basis to another general basis by doing the following

$$\mathbf{e}_{a'} = \Lambda_{a'}^{\ a}(x^k)\mathbf{e}_a \iff \mathbf{e}_a = \Lambda^{a'}_{\ a}(x^k)\mathbf{e}_{a'} \tag{1.47}$$

where

$$\Lambda_{a'}^{\ a}(x^k)\Lambda^{a'}_{\ b}(x^k) = \delta^a_b \to \mathrm{Det}(\Lambda^{a'}_{\ a}) \neq 0. \tag{1.48}$$

This is a general rule of which Eq. (1.42) is a special case. As illustrated before

$$\mathbf{X} = X^a\mathbf{e}_a = X^{a'}\mathbf{e}_{a'} \tag{1.49}$$

which in turn shows that

$$X^{a'} = \Lambda^{a'}_{\ a}X^a \iff X^a = \Lambda_{a'}^{\ a}X^{a'}. \tag{1.50}$$

This has a predictable consequence for the tensor transformation law, in such a way that

$$T^{a'b'}_{\ \ c'd'} = \Lambda^{a'}_{\ a}\Lambda^{b'}_{\ b}\Lambda^{c'}_{\ c}\Lambda^{d'}_{\ d}T^{ab}_{\ \ cd} \iff T^{ab}_{\ \ cd} = \Lambda_{a'}^{\ a}\Lambda_{b'}^{\ b}\Lambda_{c'}^{\ c}\Lambda_{d'}^{\ d}T^{a'b'}_{\ \ c'd'}. \tag{1.51}$$

This particular form is demanded in order to allow for the tensor operations, defined previously, to be well behaved for any basis. The transformation for the metric is

$$g_{a'b'} = \Lambda_{a'}^{\ a}\Lambda_{b'}^{\ b}g_{ab} \iff g_{ab} = \Lambda^{a'}_{\ a}\Lambda^{b'}_{\ b}g_{a'b'} \tag{1.52}$$

producing the invariant interval in Eq. (1.46) for displacement, in whatever basis is used. This then enables the tensor operations to be implemented as shown before and are maintained under any change of basis. An example would be to consider a unique inverse metric defined by

$$g^{ab}g_{bc} = \delta^a_c = g^a_{\ c} \tag{1.53}$$

and g_{ab} , g^{bc} can be used to lower and raise the indices defined for that base, e.g.

$$X^a = g^{ab}X_b \iff X_a = g_{ab}X^b. \tag{1.54}$$

A specific example of this would be Eq. (1.53) where it raises and lowers the indices on the metric andit follows that from Eq. (1.52) we can raise and lower indices of $\Lambda^{a'}_{\ a}$ to get

$$\Lambda_{a'}^{\ a} = g_{a'b'}\Lambda^{b'}_{\ b}g^{ab} \iff \Lambda^{a'}_{\ a} = g^{a'b'}\Lambda_{b'}^{\ b}g_{ab}. \tag{1.55}$$

The lowering and raising of an independent basis can be seen as a form of changing the basis. This amounts to taking the basis $\mathbf{e}_a$ and changing it to the dual basis $\mathbf{e}^a$ by performing the following

$$\mathbf{e}^a = g^{ab}\mathbf{e}_b, \quad \mathbf{e}_a = g_{ab}\mathbf{e}^b \iff (\mathbf{e}^a)\cdot(\mathbf{e}_b) = \delta^a_b. \tag{1.56}$$

The last equality indicates that each vector of the basis $\mathbf{e}_a$ is orthogonal to every vector of the

dual basis $\mathbf{e}_b$ except the one with the same label, with which it has a scalar product of one. This uniquely defines the geometrical relation between the two sets of basis vectors at each point of any spacetime.

In this way we can see that Eq. (1.56) is in fact a particular form of Eq. (1.47) and in a similar way Eq. (1.54) is a particular form of Eq. (1.50).

Whether or not there are coordinates for a basis $\mathbf{e}_a$, we will use a subscript comma for the directional derivative associated with a basis. When this is a coordinate basis, this is the same as $f_{,i} = \frac{\partial}{\partial x^i} f$ but more generally, if $\mathbf{e}_a = e_a{}^i \frac{\partial}{\partial x^i}$ then

$$f_{,a} = e_a(f) = e_a{}^i \frac{\partial}{\partial x^i} f \tag{1.57}$$

using the above we then define the following relation

$$Z(f) = Z^a f_{,a}.$$

1.6.2 Coordinate Basis

What we want to achieve in this section is a method or rule to determine if a basis, $\mathbf{e}_i$, is a coordinate basis or not.

This is possible by considering the commutators of the basis vectors. If the commutators produces a zero result then the basis vector used is indeed a coordinate basis, but if the result is non-zero, they are not.

First we require a working definition of a commutator $[\mathbf{X}, \mathbf{Y}]$ which is composed of two vector fields $\mathbf{X} = X^i \frac{\partial}{\partial x^i}$ and $\mathbf{Y} = Y^j \frac{\partial}{\partial y^j}$ then

$$[\mathbf{X}, \mathbf{Y}] = \mathbf{XY} - \mathbf{YX} \iff [\mathbf{X}, \mathbf{Y}]f = X(Y(f)) - Y(X(f)) \tag{1.58}$$

with the right hand equality being true for all functions $f(x^i)$. In particular if we choose $f = x^k$ and let $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$, then we find the components

$$Z^k = Z(x^k) = [X, Y](x^k) = X(Y(x^k)) - Y(X(x^k)) = X(Y^k) - Y(X^k) = X^j \frac{\partial}{\partial x^j} Y^k - Y^j \frac{\partial}{\partial x^j} X^k$$

having simplified the form. Thus when a coordinate basis is used,

$$\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \iff Z^i = Y^i_{,j} X^j - X^i_{,j} Y^j \tag{1.59}$$

gives the components of the commutator, having used Eq. (1.57).

Representing the above definition in mathematical terms leads us to the following

$$[\mathbf{e}_i, \mathbf{e}_j] = 0 \iff \exists \ \ y^i : \mathbf{e}_i = \frac{\partial}{\partial y^i}.$$

This would then mean that if we took the commutator of the basis vectors and it resulted in $[\mathbf{e}_i, \mathbf{e}_j] \neq 0$ there would then exists no coordinate for which $\mathbf{e}_i$ is a coordinate basis. We know this since in a coordinate basis, coordinates exist such that $[e_i, e_j]f = \frac{\partial}{\partial y^i}(\frac{\partial}{\partial y^j} f) - \frac{\partial}{\partial y^j}(\frac{\partial}{\partial y^i} f) = 0 \quad (f \in \mathbb{C}^2)$.

We can recall the example of an orthonormal basis $\mathbf{e}_a$ for $\mathbb{R}^2$. In this instance $[e_r, e_\theta]f = \frac{\partial}{\partial r}(\frac{1}{r}\frac{\partial f}{\partial \theta}) - \frac{1}{r}\frac{\partial}{\partial \theta}(\frac{1}{r}\frac{\partial f}{\partial r}f) = -\frac{1}{r^2}\frac{\partial f}{\partial \theta}$ for any function f, and so

$$[\mathbf{e}_r, \mathbf{e}_\theta] = -\frac{1}{r}\mathbf{e}_\theta,$$

which illustrates conclusively that $\mathbf{e}_r$ and $\mathbf{e}_\theta$ do not form a coordinate basis.

Having established the above condition for determining a coordinate basis we can conclude what the form of a transformation would be in order to preserve it as a coordinate basis. In order to achieve this, a transformation matrix of the form $\Lambda^{a'}{}_a = \frac{\partial x^{a'}}{\partial x^a}$ is required, which in turn is only possible if

$$\frac{\partial}{\partial x^b}\Lambda^{a'}{}_a = \frac{\partial}{\partial x^a}\Lambda^{a'}{}_b \iff \frac{\partial^2 x^{a'}}{\partial x^a \partial x^b} = \frac{\partial^2 x^{a'}}{\partial x^b \partial x^a}.$$

This is a specific form for a transformation and would thus not be generally applicable.

1.6.3 Tetrad Basis

A tetrad basis is a description used in which the metric components $g_{ab} = \mathbf{e}_a \cdot \mathbf{e}_b$ are constants. An orthonormal basis is a tetrad basis where

$$g_{ab} = \mathbf{e}_a \cdot \mathbf{e}_b = \text{Diag}(-1,1,1,1) =: \eta_{ab} \iff g^{ab} = \eta^{ab} = \text{Diag}(-1,1,1,1) \tag{1.60}$$

for then each component of a vector relative to this basis gives the actual magnitude of the component in that direction, and the magnitude of the vectors are given by

$$\mathbf{X} \cdot \mathbf{X} = X^a g_{ab} X^b = -(X^0)^2 + (X^1)^2 + (X^2)^2 + (X^3)^2. \tag{1.61}$$

Furthermore the invariant interval, Eq. (1.46), becomes

$$ds^2 = g_{ab}\eta^a \eta^b = -(\Lambda^0{}_i(x^k)dx^i)^2 + (\Lambda^1{}_i(x^k)dx^j)^2 + (\Lambda^2{}_i x^k dx^j)^2 + (\Lambda^3{}_i(x^k)dx^j)^2 \tag{1.62}$$

this shows the local special-relativistic structure of any arbitrary curved space-time. We can always locally choose an orthonormal basis, but it cannot be a coordinate basis unless spacetime is flat.

By choosing an orthogonal basis of vectors $(\mathbf{t}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ such that $\mathbf{t} \cdot \mathbf{t} = -1, \mathbf{x} \cdot \mathbf{x} = \mathbf{y} \cdot \mathbf{y} = \mathbf{z} \cdot \mathbf{z} = 1$ and $\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{z} = \mathbf{y} \cdot \mathbf{z} = 0$, we can see the effects of raising and lowering indices of the basis vectors. Defining $\mathbf{e}_a$ by $\{\mathbf{e}_a\} = \{\mathbf{t}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ and then $\mathbf{e}^a = g^{ab}\mathbf{e}_b$ which results in $\{-\mathbf{t}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Thus the dual basis is parallel to the original basis, but with the time vector pointing in the opposite time direction.

A Lorentz transformation preserves the orthonormality of any basis. This is the equivalent of $g_{ab} = \eta_{ab}$, $g_{a'b'} = \eta_{a'b'}$ and therefore it will be a Lorentz transformation if

$$\eta_{a'b'} = \Lambda_{a'}{}^a \Lambda_{b'}{}^b \eta_{ab}, \tag{1.63}$$

which ensures that the equations (1.60) to (1.62) are true in both the old and new frame. This set of transformations forms a group called the Lorentz group.

These transformations can be composed of spatial rotations and 'boosts' or velocity transformations, such as

$$\Lambda^{i}{}_{i'}(v) = \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.64}$$

with $\gamma(v) = (1 - v^2)^{-1/2}$ and $c = 1$. This can be verified by taking the matrix form of $\eta_{a'b'} = (\Lambda_{a'}{}^{a})(\eta_{ab})(\Lambda_{b'}{}^{b})^T$ where the left hand index denotes a row and the right hand index corresponds to a column, while the transpose of the right hand matrix is taken in order for the summation convention to adhere to the rules of matrix multiplication,

$$\eta_{a'b'} = \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As a result of the group property a Lorentz transformation generates an associated family of transformations that preserves the Minkowski metric form. There do, however, exist some (non-unique) transformation $\lambda_{a}{}^{i}(x^k)$ that locally transforms an arbitrary metric g_{ij} to the orthonormal form where we will have

$$\eta_{ab} = \lambda_{a}{}^{i}(x^k)\lambda_{b}{}^{j}(x^k)g_{ij}(x^k) \iff g_{ij}(x^k) = \lambda^{a}{}_{j}(x^k)\lambda^{b}_{j}\eta_{ab}. \tag{1.65}$$

Then the transformation

$$A_{j}{}^{i}(x^k) = \lambda^{a}{}_{j}(x^k)\Lambda_{a}{}^{b}(x^k)\lambda_{b}{}^{i}(x^k) \tag{1.66}$$

will preserve the form of g_{ij} if $\Lambda^{a}{}_{b}(x^k)$ is a Lorentz transformation.

Chapter 2
Gravity

This chapter aims to first introduce the idea that gravitational effects can be determined entirely from the relative paths of free falling observers.

If we consider the path taken by a free falling observer through time and space the result is a curve that is labelled by measuring it's initial velocity, position and time at each point in space with a "good clock" or rather a set of time coordinates for a local inertial frame. The curves specifically considered here are geodesics and the parameterisation (labelling) of the geodesics is called an affine parameter, λ. The parameterisation is not specific about the choice of origin and units or scaling and is thus unique up to a linear transformation, $\lambda_{\text{new}} = a\lambda_{\text{old}} + b$.

Knowing that the physics does not change along this path and that locally all the laws of physics are well defined and preserved we must have a conceptual appreciation of how things do change from position to position along this curve. For this we require the covariant derivative.

It will then be shown how the relative change in the geodesics account for gravity or, put another way, curvature of the spacetime. This information is captured in the Riemann curvature tensor. This perspective is then used to show how it applies to Newtonian gravity. After the introduction of Einstein's field equations the Newtonian limit is shown and a table illustrates the fundamental differences between Newtonian gravity and Einstein's theory of gravity. The chapter concludes with a review of some common and useful exact solutions for the Einstein field equations.

2.1 Covariant Derivative

Let us consider a curved manifold, M, with a vector field, $\mathbf{v}$, defined on it. Since we are dealing with a varying vector field each point on the manifold will have a different vector associated with it. If we want to characterise the variation in the vector field from point to point on the manifold we should start by first considering two points on the manifold and comparing the vectors associated with them.

Each point has an associated tangent space and basis vectors, this is equivalent to considering each neighbourhood of a point being locally flat or Euclidean. In order to evaluate two vectors at differing points we have to devise a means of comparing the two associated tangent spaces with each other.

We can do this by using parallel transport. We first connect the two points under consideration by a curve. The curve is parameterised by λ. A measure of change from the first to the second point is done by evaluating the difference between the two respective vectors given at each point. This can be seen as shifting the first vector along the tangent of the adjoining curve in steps of λ in a parallel fashion, without changing its original orientation, to the point where the two vectors can be subtracted to establish the deviation. This is illustrated in Figure 2.1.

This is a change in $\mathbf{v}$ with respect to the tangent vector $\mathbf{u}$ of a curve, represented as $\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v}$. It

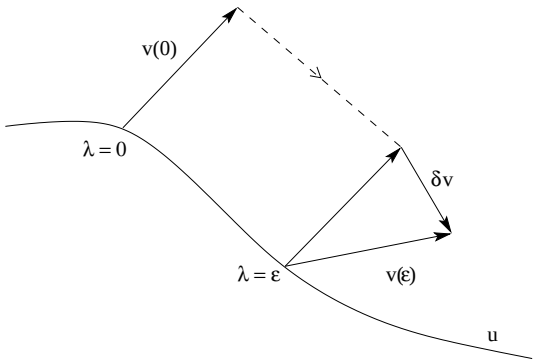


Figure 2.1: An illustration of the mechanics of the covariant derivative given in Eq. 2.1. The vector v is parallel transported along the curve u from $\lambda = 0$ to $\lambda = \varepsilon$. δv is the result of the covariant derivative. Based on a diagram from [MTW73].

is essentially a standard definition of a derivative for the vectors along a curve

$$\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} = \lim_{\epsilon \longrightarrow 0}\left[\frac{\mathbf{v}(\lambda_0 + \epsilon)_{\ \parallel \text{ transported to } \lambda_0} - \mathbf{v}(\lambda_0)}{\epsilon}\right]. \tag{2.1}$$

To better understand the notion of a covariant derivative it is useful to consider Eq. (2.1) in component form. The complication that arises in component form is that we have to take into account the fact that each basis vectors for each of the respective tangent spaces can be chosen completely arbitrarily. In an Euclidean or flat space we could arrange to have only one set of basis vectors for the entire space (this would mean that we only require the derivative above to calculate the change in the vector field) but if we would change the basis from point to point we would get a perceived deviation. We must therefore take into account the deviation caused by the change in basis from point to point in each of the tangent spaces and then correct for it.

We can define the change in the basis relative to each other as a connection. This is a way of keeping account of how the basis changes as we move from tangent space to tangent space. The connection will represent a twisting and turning of the basis along the curve. This can be represented in terms of the derivative we have already discussed by asking the question how do the basis vectors $\mathbf{e}_\alpha$ change along the curve with tangent vector $\mathbf{e}^\beta$:

$$\boldsymbol{\nabla}_\beta \mathbf{e}_\alpha = \mathbf{e}_\nu \Gamma^\nu{}_{\alpha\beta} \quad \text{where} \quad \boldsymbol{\nabla}_\beta = \boldsymbol{\nabla}_{\mathbf{e}^\beta}. \tag{2.2}$$

Since the covariant derivative is itself a vector the connection coefficients $\Gamma^\nu{}_{\alpha\beta}$ can be seen as the expansion coefficients of the covariant derivative in terms of the basis vectors. The connection coefficients are not tensors and thus do not transform as such.

An important point to raise is that the connection coefficients can be calculated from the metric of the manifold. This can be explicitly derived, [MTW73], This allows us an alternate means to determine what the connection cofficients are by only using the metric,

$$\Gamma^\alpha{}_{\beta\gamma} = g^{\alpha\mu}\Gamma_{\mu\beta\gamma}, \tag{2.3}$$

where,

$$\begin{aligned}\Gamma_{\mu\beta\gamma} &= \frac{1}{2}(g_{\mu\beta,\gamma} + g_{\mu\gamma,\beta} - g_{\beta\gamma,\mu}) \\ &= \frac{1}{2}\left(\frac{\partial g_{\mu\beta}}{\partial x^\gamma} + \frac{\partial g_{\mu\gamma}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\mu}\right). \end{aligned} \tag{2.4}$$

The above equation employed the notational shorthand where a partial derivative is given by a comma.

The covariant derivative is the same as an ordinary derivative while also combining the attributes that the connection introduces. This allows us to solve the problem of the perceived deviation as the basis changes from point to point. The basis can be arbitrarily chosen or altered and this will not impact on the covariant derivative. Put another way the covariant derivative allows the basis to be arbitrarily transformed but the result will remain the same. In component form this means that the covariant derivative transforms covariantly under an arbitrary local change of basis.

The Eq. (2.5) indicates how the covariant derivative looks once it has worked in on a vector in component form.

$$\begin{aligned}\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} = \boldsymbol{\nabla}_{\mathbf{u}} v^\beta e_\beta &= \frac{\partial v^\beta}{\partial \lambda} e_\beta + v^\beta \boldsymbol{\nabla}_{\mathbf{u}} e_\beta \\ &= \underbrace{\frac{\partial v^\beta}{\partial \lambda} e_\beta}_{\text{change with respect to curve}} + \underbrace{v^\beta \, \Gamma^\nu{}_{\beta\mathbf{u}} e_\nu}_{\text{change with respect to basis}} . \end{aligned} \tag{2.5}$$

This gives a clear feel for the two aspects the covariant derivative tries to account for.

The covariant derivative has a set of properties associated with it that are often used to define it. They are summarised in [MTW73] as being:

- Symmetry: $\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} - \boldsymbol{\nabla}_{\mathbf{v}}\mathbf{u} = [\boldsymbol{\nabla}_{\mathbf{u}}, \boldsymbol{\nabla}_{\mathbf{v}}] = [\mathbf{u}, \mathbf{v}]$ for any vector field $\mathbf{u}$ and $\mathbf{v}$, where $[,]$ are the commutator brackets.
- Chain rule: $\boldsymbol{\nabla}_{\mathbf{u}}(f\mathbf{v}) = f\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} + \mathbf{v}\partial_{\mathbf{u}} f$ for any function f, vector field $\mathbf{v}$, and vector $\mathbf{u}$, where ∂_u is a functional derivative.
- Additivity 1: $\boldsymbol{\nabla}_{\mathbf{u}}(\mathbf{v} + \mathbf{w}) = \boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} + \boldsymbol{\nabla}_{\mathbf{u}}\mathbf{w}$ for any vector fields $\mathbf{v}$ and $\mathbf{w}$, and vector $\mathbf{u}$.
- Additivity 2: $\boldsymbol{\nabla}_{f\mathbf{u}+h\mathbf{n}}\mathbf{v} = f\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{v} + h\boldsymbol{\nabla}_{\mathbf{n}}\mathbf{v}$ for any vector field $\mathbf{v}$, vector fields $\mathbf{u}$ and $\mathbf{n}$, and numbers or functions f and h.

2.2 Example for two dimensional polar coordinates

To illustrate the nature of the connection we will work through an example of a 2-dimensional polar coordinate system with basis vectors, $\mathbf{e}_r$ and $\mathbf{e}_\theta$. We will use a transformation from the more familiar Euclidean coordinates to polar coordinates.

From geometric considerations, Figure 2.2, we can conclude that there will be no relative change along the tangent vector of $\mathbf{e}_r$ since this is a basis vector being parallel transported along

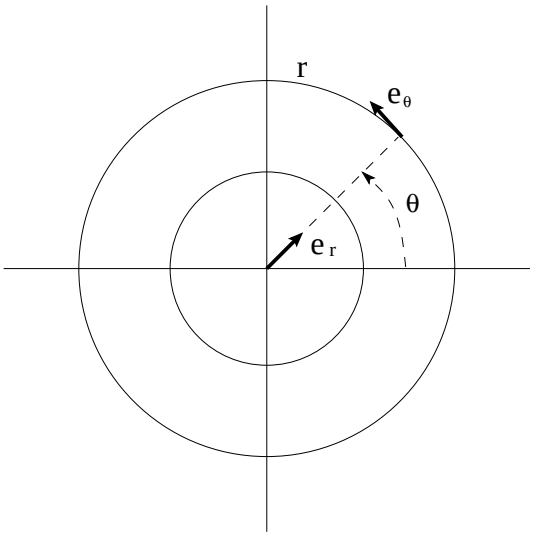


Figure 2.2: A diagram showing polar coordinates in two dimensions. The basis vectors $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are shown along with the labelling of the radius r and the angle θ. It can be seen that if $\mathbf{e}_r$ or $\mathbf{e}_\theta$ are parallel transported along $\mathbf{e}_r$ there will be no deviation and thus the covariant derivative will be zero. Only along parallel transported vectors along $\mathbf{e}_\theta$ will produce a deviation and thus a positive result.

a straight line, it is clear then that

$$\boldsymbol{\nabla}_r \mathbf{e}_r = 0, \quad \boldsymbol{\nabla}_r \mathbf{e}_\theta = 0 \quad \text{while} \quad \boldsymbol{\nabla}_\theta \mathbf{e}_r = a\mathbf{e}_\theta, \quad \boldsymbol{\nabla}_\theta \mathbf{e}_\theta = b\mathbf{e}_r, \tag{2.6}$$

for some function a, b. We transform $\mathbf{e}_r$ and $\mathbf{e}_\theta$ into their Euclidean form,

$$\mathbf{e}_r = \cos\theta \mathbf{e}_x + \sin\theta \mathbf{e}_y, \quad \mathbf{e}_\theta = -r\sin\theta \mathbf{e}_x + r\cos\theta \mathbf{e}_y. \tag{2.7}$$

Then, for example, we could consider

$$\begin{aligned} \boldsymbol{\nabla}_\theta \mathbf{e}_r &= \boldsymbol{\nabla}_\theta(\cos\theta \mathbf{e}_x + \sin\theta \mathbf{e}_y) \\ &= \boldsymbol{\nabla}_\theta \cos\theta \mathbf{e}_x + \boldsymbol{\nabla}_\theta \sin\theta \mathbf{e}_y \end{aligned}$$

noting that $\boldsymbol{\nabla}_\theta = \boldsymbol{\nabla}_{\mathbf{e}_\theta} = \boldsymbol{\nabla}_{(-r\sin\theta \mathbf{e}_x + r\cos\theta \mathbf{e}_y)}$ from Eq. (2.7), we can then apply the second additive property of the covariant derivative,

$$= (-r\sin\theta \boldsymbol{\nabla}_x + r\cos\theta \boldsymbol{\nabla}_y)\cos\theta \mathbf{e}_x + (-r\sin\theta \boldsymbol{\nabla}_x + r\cos\theta \boldsymbol{\nabla}_y)\sin\theta \mathbf{e}_y$$

using $\boldsymbol{\nabla}_x = \frac{\partial r}{\partial x}\boldsymbol{\nabla}_r + \frac{\partial \theta}{\partial x}\boldsymbol{\nabla}_\theta = \cos\theta \boldsymbol{\nabla}_r - \frac{\sin\theta}{r}\boldsymbol{\nabla}_\theta$ where $\frac{\partial r}{\partial x} = \cos\theta$ and $\frac{\partial \theta}{\partial x} = \frac{\partial}{\partial x}(\tan^{-1}(\frac{y}{x}) = -\frac{y}{x^2+y^2} = -\frac{y}{r}\frac{1}{r} = -\frac{\sin\theta}{r}$ produces,

$$\begin{aligned} &= r\left(-\sin\theta\left(\frac{\sin^2\theta}{r}\right)\mathbf{e}_x - \cos\theta\left(\frac{\cos\theta\sin\theta}{r}\right)\mathbf{e}_x + \sin\theta\left(\frac{\cos\theta\sin\theta}{r}\right)\mathbf{e}_y + \cos\theta\left(\frac{\cos^2\theta}{r}\right)\mathbf{e}_y\right) \\ &= [-\sin\theta \mathbf{e}_x + \cos\theta \mathbf{e}_y] \\ &= \frac{1}{r}\mathbf{e}_\theta, \end{aligned}$$

where we used Eq. (2.7). Similarly we can calculate

$$\begin{aligned}
\boldsymbol{\nabla}_\theta \mathbf{e}_\theta &= \boldsymbol{\nabla}_\theta(-\sin\theta\mathbf{e}_x + \cos\theta\mathbf{e}_y) \\
&= \boldsymbol{\nabla}_\theta(-\sin\theta\mathbf{e}_x) + \boldsymbol{\nabla}_\theta\cos\theta\mathbf{e}_r \\
&= -\sin\theta\boldsymbol{\nabla}_x(-\sin\theta)\mathbf{e}_x + \cos\theta\boldsymbol{\nabla}_y(-\sin\theta)\mathbf{e}_x - \sin\theta\boldsymbol{\nabla}_x\cos\theta\mathbf{e}_y + \cos\theta\boldsymbol{\nabla}_y\cos\theta\mathbf{e}_y \\
&= -\sin\theta\left(\frac{\cos\theta\sin\theta}{r}\right)\mathbf{e}_x + \cos\theta\frac{-\cos^2\theta}{r}\mathbf{e}_x - \sin\theta\left(\frac{\sin^2\theta}{r}\right)\mathbf{e}_y + \cos\theta\left(\frac{-\cos\theta\sin\theta}{r}\right)\mathbf{e}_y \\
&= -\frac{1}{r}[\cos\theta\mathbf{e}_x + \sin\theta\mathbf{e}_y] \\
&= -\frac{1}{r}\mathbf{e}_r
\end{aligned}$$

Since we already know that $\frac{\partial \mathbf{e}_\alpha}{\partial x^\beta} = \Gamma^\mu{}_{\alpha\beta}\mathbf{e}_\mu$ from Eq. (2.6), we can thus with little effort calculate the connection coefficients from the above result,

$$\Gamma^\theta{}_{\theta r} = \frac{1}{r}, \quad \Gamma^r{}_{\theta\theta} = -\frac{1}{r},$$

these being the only non-zero results.

2.3 Geodesics

Having established what the covariant derivative is, we can define a geodesic as a vector, $\mathbf{u}$, that is being parallel transported along the tangent vector $\mathbf{u}$ without any deviation,

$$\boldsymbol{\nabla}_\mathbf{u}\mathbf{u} = 0. \tag{2.8}$$

In a flat space where a tangent vector is parallel transported along itself without any deviation a straight line is produced. A geodesic in a flat space is thus a straight line. In a curved space we can still get geodesics but it isn't necessarily a straight line but rather a locally straight line in the given neighbourhood of the tangent vector. This is a very powerful concept that illustrates some of the local nature of geodesics that is important to General Relativity and will be touched on again in more detail later.

In order to place a geodesic equation in the form of a differential equation that will often be used, we require the definition of a geodesic, Eq (2.8), and then write this in component form,

$$\begin{aligned}
0 = \mathbf{u}^\alpha_{;\beta} &= (\mathbf{u}^\alpha_{,\beta} + \Gamma^\alpha{}_{\gamma\beta}\mathbf{u}^\gamma)\mathbf{u}^\beta \\
&= \frac{\partial}{\partial x^\beta}\left(\frac{dx^\alpha}{d\lambda}\right)\frac{dx^\beta}{d\lambda} + \Gamma^\alpha{}_{\gamma\beta}\frac{dx^\gamma}{d\lambda}\frac{dx^\beta}{d\lambda},
\end{aligned}$$

which reduces to the desired form

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\gamma\beta}\frac{dx^\gamma}{d\lambda}\frac{dx^\beta}{d\lambda} = 0, \tag{2.9}$$

where, λ, is the affine parameter and the notational shorthand was used above where a semi-colon represents the covariant derivative.

An acknowledged property of straight lines is that they are the shortest distance between two

points in a flat space. This is a characteristic of geodesics that extends beyond flat space and can be used as a defining attribute that applies in curved space as well. To examine this property and show that this leads to some properties familiar to us in Special Relativity , we can use the action $\phi = \int_P^Q L(x^i, \dot{x}^j)dv$ which gives us the local extremal value (maxima or minima) on a space of curves $x^i(v)$ joining P to Q if and only if

$$\frac{\partial L}{\partial x^s} - \frac{d}{dv}\Big(\frac{\partial L}{\partial \dot{x}^s}\Big) = 0. \tag{2.10}$$

Consider a Lagrangian , $L(x^i, \dot{x}^j) = F(S)$, where we define $S(x^i, \dot{x}^j) := g_{ij}(x^k)\dot{x}^i\dot{x}^j$, resulting in $\frac{dF}{dS} \neq 0$. Calculating

$$\frac{\partial S}{\partial x^s} = g_{ij,s}\dot{x}^i\dot{x}^j, \quad \frac{\partial S}{\partial \dot{x}^s} = g_{sj}\dot{x}^j + g_{is}\dot{x}^i = 2g_{is}\dot{x}^i$$

where we have used the fact that the metric, g_{ij}, is symmetric and then substituting this result into Eq. (2.10),

$$\begin{aligned}\frac{dF}{dS}\frac{\partial S}{\partial x^s} - \frac{d}{dv}\Big(\frac{dF}{dS}\frac{\partial S}{\partial \dot{x}^s}\Big) &= \frac{dF}{dS}g_{ij,s}\dot{x}^i\dot{x}^j - \frac{d}{dv}\Big(\frac{dF}{dS}2g_{is}\dot{x}^i\Big) = 0 \\ \frac{dF}{dS}g_{ij,s}\dot{x}^i\dot{x}^j - \frac{d^2F}{dS^2}\frac{dS}{dv}2g_{is}\dot{x}^i - \frac{dF}{dS}(2g_{is,j}\dot{x}^i\dot{x}^j + 2g_{is}\ddot{x}^i) &= 0.\end{aligned}$$

Then dividing by $(-2\frac{dF}{dS})$ gives,

$$g_{is}\ddot{x}^i + \frac{1}{2}(g_{is,j} + g_{js,i} - g_{ij,s})\dot{x}^i\dot{x}^j = G(S)g_{is}\dot{x}^i \quad \text{where} \quad G(S) = -\Big(\frac{d^2F}{dS^2}\frac{dF}{dS}\Big)\frac{dS}{dv}.$$

Now multiply by g^{sk} produces,

$$\ddot{x}^k + \Gamma^k{}_{ij}\dot{x}^i\dot{x}^j = G(S)\dot{x}^k. \tag{2.11}$$

If we choose a spacelike curve, with proper distance s, then $s = \int (S)^{1/2}dv$, so $\frac{ds}{dv} = (S)^{1/2}$ and $\frac{d^2s}{dv^2} = \frac{1}{2}(S)^{-1/2}\frac{dS}{dv}$. Hence

$$s = av + b \iff \frac{d^2s}{dv^2} = 0 \iff \frac{dS}{dv} = 0 \implies G(S) = 0$$

where a, b are constants (recovering the affine transformation). If we in particular choose $v = s$ then we have an affinely parameterised geodesic as in Eq. (2.9), for all $F(S)$. This then shows that by choosing $F(S) = S^{1/2}$ a curve that is locally *minimising* proper distance is a geodesic, with proper distance an affine parameter (we know from the length contraction formula in SR that this should always be a minimum). Similarly in the timelike case, if we choose $F(S) = (-S)^{1/2}$, it can be shown that a curve locally *maximising* proper time is a geodesic, with proper time being the affine parameter. The fact that it is now being *maximised* for the timelike curve as opposed to *minimised* in the spacelike case is because of the minus sign introduced by virtue of the Lorentzian signature of the metric.

2.4 Relative Geodesics

A geodesic can be seen as an equation of motion of a freely moving neutral test particle. This is based on the plausible assumption that a freely moving neutral particle always takes a unique minimised path between two points.

If we now consider the fact that gravity causes a spacetime curvature then we would see this manifest in the relative motion of freely falling neutral test particles or put another way, the relative change in neighbouring geodesics. This is what we are going to explore in the next section and try and establish the nature of relative geodesics and how this can then be interpreted as a spacetime curvature which accounts for the effects of gravity.

We have a family of geodesics and a tangent vector, $\mathbf{u} = \frac{\partial}{\partial \lambda}$, defined along it. We introduce a separation vector, $\mathbf{n} = \frac{\partial}{\partial \mathbf{n}}$, between the adjacent geodesics, separated by Δn.

If we consider the change of the covariant derivative of the separation vector, $\mathbf{n}$, and the tangent vector, $\mathbf{u}$, we arrive at the relative acceleration of the geodesics.

By taking the covariant derivative of the separation vector along the geodesic we get a result telling us how much the separation vector varies from point to point along the geodesic thus enabling us to know how the neighbouring geodesic changes along λ. If the separation vector was to be parallel transported along the geodesic it would imply that the covariant derivative is zero and there is no change between the geodesics.

We ultimately want to know what the relative relationship is between the neighbouring geodesics and therefore need the rate of change of the covariant derivative along the geodesic. This produces the relationship for relative acceleration of neighbouring geodesics

$$\boldsymbol{\nabla}_{\mathbf{u}} \boldsymbol{\nabla}_{\mathbf{u}} \mathbf{n}. \tag{2.12}$$

Figure 2.3 illustrates in more detail what is meant by relative acceleration.

If we consider a family of geodesics, Figure 2.4, and we introduce a separation vector, $\mathbf{n} = \frac{\partial}{\partial n}$, between the adjacent geodesics, then we can define the relative change in geodesics as the deviation of the geodesic, $\boldsymbol{\nabla}_{\mathbf{u}} \mathbf{u}$, with respect to the separation vector, $\mathbf{n}$. This is written as $\boldsymbol{\nabla}_{\mathbf{n}} \boldsymbol{\nabla}_{\mathbf{u}} \mathbf{u}$.

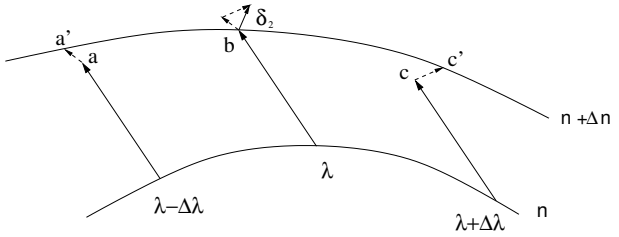


Figure 2.3: This diagram shows a parallel transported separation vector, u , which outlines what a constant separation between the two geodesics would have been. The covariant derivative, $\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{n}$, at $\lambda - \frac{1}{2}\Delta\lambda$ and $\lambda + \frac{1}{2}\Delta\lambda$, is $-aa'/\Delta\lambda\Delta n$ and $cc'/\Delta\lambda\Delta n$. The second derivative, $\boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{n}$, at λ is thus $\frac{(\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{n})_{\lambda+\frac{1}{2}\Delta\lambda} - (\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{n})_{\lambda-\frac{1}{2}\Delta\lambda}}{\Delta\lambda}$ which reduces to $\delta_2/((\Delta\lambda)^2\Delta n)$. Based on diagram from [MTW73].

Imagine that we parallel transport a geodesic along the separation vector. Doing this would be equivalent to saying that we have a family of geodesics that are identical and are evenly

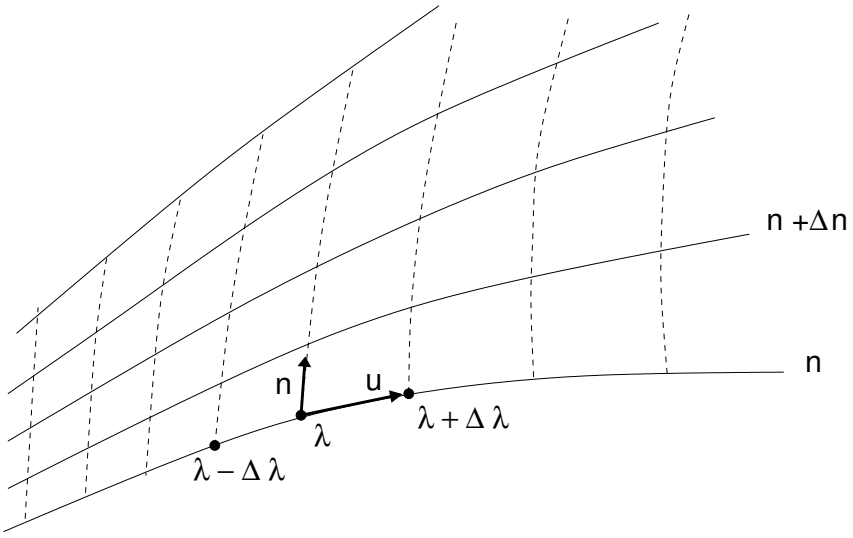


Figure 2.4: A diagram of a family of geodesics with affine parameter λ, a geodesic selector parameter n which identifies the geodesic and the two respective tangent vectors n and u. Based on diagram from [MTW73].

spaced from each other by the separation vector. The equation for this is introduced below and re-arranged to the following form,

$$\begin{aligned} \boldsymbol{\nabla}_{\mathbf{n}}[\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{u}] &= 0 \\ \boldsymbol{\nabla}_{\mathbf{n}}\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{u} + (-\boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{n}}\mathbf{u} + \boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{n}}\mathbf{u}) &= 0 \end{aligned}$$

where $\boldsymbol{\nabla}_{\mathbf{n}}\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{u} - \boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{n}}\mathbf{u} = \big[\boldsymbol{\nabla}_{\mathbf{n}}, \boldsymbol{\nabla}_{\mathbf{u}}\big]$, implies that we can simplify to

$$\boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{n}}\mathbf{u} + \big[\boldsymbol{\nabla}_{\mathbf{n}}, \boldsymbol{\nabla}_{\mathbf{u}}\big]\mathbf{u} = 0,$$

using the Symmetry property of the covariant derivative in Section 2.1 we achieve

$$\boldsymbol{\nabla}_{\mathbf{u}}\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{n} + [\boldsymbol{\nabla}_{\mathbf{n}}, \boldsymbol{\nabla}_{\mathbf{u}}]\mathbf{u} = 0.$$

This produces two distinct contributing terms. The first we are already familiar with and recognise as being the term for relative acceleration of geodesics. The second term is called the Riemann curvature tensor (still to be defined below). This term balances or counteracts the relative acceleration of the geodesics to produce a zero result. It is thus the term in which the effects of gravity manifest and is analogous to the "tide producing gravitational force" in Newtonian gravity, [MTW73].

The Riemann tensor is very significant when considering gravity but isn't complete in its current formulation and requires some refinement in order to make it more generic. Consider the vector field C_{new} and C_{old} related to each other in the following way $C_{\text{new}}(P) = f(P)C_{\text{old}}(P)$ where f is an arbitrary function and $f(P_0) = 1$. In the following example we show that

$[\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{new}} \neq [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}}$, at P_0, by calculating,

$$\begin{aligned}
[\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{new}} - [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}} &= f(P)[\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}}(P) + C_{\mathrm{old}}(P)[\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]f(P) - [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}} \\
&= f(P)[\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}}(P) + C_{\mathrm{old}}(P)\boldsymbol{\nabla}_{[\mathrm{A,B}]}f(P) - [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}} \\
&= [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}}(P_0) + C_{\mathrm{old}}(P_0)\boldsymbol{\nabla}_{[\mathrm{A,B}]}f(P_0) - [\boldsymbol{\nabla}_{\mathrm{A}}, \boldsymbol{\nabla}_{\mathrm{B}}]C_{\mathrm{old}} \\
&= C_{\mathrm{old}}\boldsymbol{\nabla}_{[\mathrm{A,B}]}.
\end{aligned}$$

This difference can simply be rectified by subtracting the problematic term. The new and complete form of the Riemann curvature tensor is thus given by

$$\mathcal{R}(\mathbf{u}, \mathbf{n}) = [\boldsymbol{\nabla}_{\mathbf{u}}, \boldsymbol{\nabla}_{\mathbf{n}}] - \boldsymbol{\nabla}_{[\mathbf{u},\mathbf{n}]}. \tag{2.13}$$

The Riemann curvature tensor can be expressed in coordinate form, $\mathcal{R}^{\alpha}{}_{\beta\gamma\delta} = \frac{\partial \Gamma^{\alpha}{}_{\beta\delta}}{\partial x^{\gamma}} - \frac{\partial \Gamma^{\alpha}{}_{\beta\gamma}}{\partial x^{\delta}} + \Gamma^{\alpha}{}_{\mu\gamma}\Gamma^{\mu}{}_{\beta\delta} - \Gamma^{\alpha}{}_{\mu\delta}\Gamma^{\mu}{}_{\beta\gamma}$. In a flat space the connection coefficients are zero, Section 2.1. This implies that the Riemann curvature tensor is zero as well, which can be seen from the above equation.. A locally flat or inertial frame of coordinates can be derived by setting the connection coefficients to zero. This realisation is known as the Riemann normal coordinate system.

The Riemann curvature tensor can be contracted to form the Ricci curvature tensor which is often used in general relativity,

$$\mathcal{R}_{\alpha\beta} \equiv \mathcal{R}^{\mu}{}_{\alpha\mu\beta}. \tag{2.14}$$

2.5 Geometric interpretation of Newtonian Gravitation

In Newtonian gravity we must discard our concept of metrics, the limited speed of light and proper time. In Newtonian gravity we only have a *universal time*. The idea, however, that the equivalence principle is exclusive to general relativity is incorrect and it can be applied to Newtonian gravity in a straight forward way to deliver a geometric interpretation of Newtonian gravity. This was explored and illustrated by Cartan, [Car23] and [Car24].

When we think of a Newtonian picture we imagine a curved trajectories of neutral test particles on an uninteresting Euclidean background. Cartan showed that if one abandons this idea and instead views the trajectories as geodesics we arrive at a curved spacetime. This is the same as considering curved lines that become straight and thus force the background to become curved.

We start with Newton's equation of motion

$$\begin{aligned}
F &= m\mathbf{a}, \\
\mathbf{a} &= -\boldsymbol{\nabla}\phi, \\
\frac{d^2x^j}{dt^2} + \frac{\partial \phi}{\partial x^j} &= 0,
\end{aligned}$$

where $\phi = GM/r$ is the Newtonian potential. This result describes curved paths in Euclidean space. Remembering that we are dealing with universal time, the test particles clock reflects this and is equal to the universal time or some multiple there of, $\tau = at + b$. Using this fact we have

that $\frac{d^2t}{d\tau^2} = 0$ which can be substituted into the above equation to produce

$$\frac{d^2x^i}{d\tau^2} + \frac{\partial\phi}{\partial x^j}\left(\frac{dt}{d\tau}\right)^2 = 0. \tag{2.15}$$

This is the geodesics equation for Newtonian gravity. If we compare it to the geodesic equation we derived in the previous section, Eq. (2.9), we can read off the Γ terms,

$$\Gamma^j{}_{00} = \frac{\partial\phi}{\partial x^j}, \tag{2.16}$$

which is the only non-zero component of the Γ's. Inserting this into the curvature tensor, Eq. (2.13),

$$\mathcal{R}^j{}_{0k0} = -\mathcal{R}^j{}_{00k} = \frac{\partial^2\phi}{\partial x^j\partial x^k} \tag{2.17}$$

contracting this to produce the Ricci tensor, Eq. (2.14),

$$\mathcal{R}_{00} = \mathcal{R}^j{}_{0j0} = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = \boldsymbol{\nabla}^2\phi. \tag{2.18}$$

This is the Laplace of gravitational potential and can thus then be written in terms of the curvature tensor,

$$\mathcal{R}_{00} = \mathcal{R}^j{}_{0j0} = 4\pi G\rho. \tag{2.19}$$

This contains all the information of Newtonian gravity. It is insightful to derive the relative acceleration of a test particle within this framework of Newtonian geodesics.

Differentiate Eq. (2.15) with respect to the parameter s, noting that $\frac{\partial}{\partial s} = n^i\frac{\partial}{\partial x^i}$.

$$\frac{\partial^2 n^i}{\partial t^2} + n^i\frac{\partial^2\phi}{\partial x^i\partial x^j} = 0. \tag{2.20}$$

This is equivalent to Eq. (2.13), where one can recognise the second term as the tide producing (or the curvature) term.

2.5.1 Key points to geometric picture of Newtonian gravitational theory

The Newtonian version of gravity has a universal time and an absolute space. Combining space and time allows for a version of curved Newtonian spacetime. The geometry of each space slice is completely flat. Parallel transporting a vector around a closed curve lying entirely in a space slice will result in it returning to its starting point unchanged, while transporting it forward in time around the same closed path will produce a non zero result. Geodesics of a space slice that are initially parallel always remain parallel, but in spacetime geodesics initially parallel get pried apart or pushed together by spacetime curvature,

$$\boldsymbol{\nabla}_{\mathbf{u}}\nabla_{\mathbf{u}}\mathbf{n} + \mathcal{R}(\mathbf{u},\mathbf{n})\mathbf{u} = 0.$$

or equivalently in Galilean coordinates Eq. (2.20).

- There exists a universal time t, a set of Cartesian coordinates x^j and a Newtonian gravi-

tational potential Φ.

- The density of mass ρ generates the Newtonian potential by Poisson's equation,

$$\nabla^2\phi \equiv \frac{\partial^2\phi}{\partial x^j \partial x^j} = 4\pi\rho.$$

- The equation for motion of a freely falling particle is

$$\frac{d^2x^j}{dt^2} + \frac{\partial\phi}{\partial x^j} = 0.$$

- 'Ideal rods' measure the coordinate lengths while 'ideal clocks' measure universal time.

2.6 Einsteins Field Equations

Einsteins idea to explain gravity focused on the conjecture that matter determines spacetime structure. The metric is central to this idea and encompasses all of the elements of the above conjecture.

Since Newtonian gravity is a very successful theory we wouldn't expect that a new theory would be too far off the mark with respect to its structure. For this reason we would accept that when we consider a coordinate description of spacetime we would find a second order equation for the metric components that encodes the gravitational field. This would allow us to retrieve the second order equation in the Newtonian limit.

Taking the Christoffel relation

$$\Gamma^a{}_{ed} = g^{ac}\Gamma_{ced}, \qquad \Gamma_{ced} = \frac{1}{2}\{g_{cd,e} + g_{ec,d} - g_{de,c}\}, \tag{2.21}$$

and then the curvature tensor

$$\mathcal{R}_f{}^a{}_{bc} = \Gamma^a{}_{bf,c} - \Gamma^a{}_{cf,b} + \Gamma^e{}_{bf}\Gamma^a{}_{ce} - \Gamma^e{}_{cf}\Gamma^a{}_{be}, \tag{2.22}$$

remembering that we use the shorthand notation, $\Gamma^a{}_{bf,c} = \partial\Gamma^a{}_{bf}/\partial x^c$,results in a second order equation in terms of the metric tensor components. This is a positive result in trying to establish that geometric gravity is consistent with the notion that matter causes spacetime curvature.

Timelike geodesics represent freely falling objects. The relative geodesics are governed by a second order equation therefore the tidal force (Newton) of gravity is represented by curvature and so should be determined by the distribution of matter in spacetime through a suitable equation. This equation should set the relationship between matter and curvature.

The energy-momentum tensor, T_{ab}, characterises all types of matter irrespective of the form

$$T_{ab} = T_{(ab)} \qquad T^{ab}{}_{;b} = 0$$

which is symmetric and conserved. This tensor has only two indices and can not naturally be equated to the curvature tensor which has four indices, however the Ricci tensor offers a solution

with two indices

$$\mathcal{R}_{ab} = \mathcal{R}^{s}{}_{asb} = \Gamma^{s}{}_{ba,s} - \Gamma^{s}{}_{sa,b} + \Gamma^{e}{}_{ba}\Gamma^{s}{}_{se} - \Gamma^{e}{}_{sa}\Gamma^{s}{}_{be} \tag{2.23}$$

where $\mathcal{R}_{ab} = \mathcal{R}_{(ab)}$ is also symmetric. A plausible proposal would thus be

$$\mathcal{R}_{ab} = \kappa T_{ab} \tag{2.24}$$

where κ is a coupling constant. This was Einsteins initial proposal but failed due to the Bianchi identity. This is evident when we take the divergence of the proposed equation

$$\mathcal{R}^{ab}{}_{;b} = 0 \quad \Longrightarrow \mathcal{R}_{,a} = 0 \quad \Longleftrightarrow \quad \mathcal{R} = \text{const}$$

where $\mathcal{R} := \mathcal{R}^{a}{}_{a}$ is called the curvature scalar. The above equation results in $\mathcal{R} = \kappa T^{a}{}_{a}$ from Eq. (2.24) which in turn requires a constant trace of the energy-momentum tensor. This is far too restrictive and results in instances that are unphysical.

By taking into account the form of the Bianchi identity another solution arises in the form

$$G_{ab} = \mathcal{R}_{ab} - \frac{1}{2}\mathcal{R}g_{ab} \Longrightarrow G_{ab} = G_{(ab)} \tag{2.25}$$

where $G_{ab} = G_{(ab)}$ is symmetric and satisfies the Bianchi identity, $G^{ab}{}_{;b} = 0$. We thus arrive at a refined version of Einsteins initial equation, replacing Eq. (2.24) with

$$G_{ab} = \kappa T_{ab}.$$

This equation possesses the prospect of satisfying all the conditions required in order to establish a relationship between matter and spacetime curvature. The only problem that appears is that another solution also exists

$$G_{ab} + \Lambda g_{ab} = \kappa T_{ab}.$$

For the purposes of the rest of this discussion we will employ Occam's razor where the simplest solution of $\Lambda = 0$ is assumed. This then produces the final Einstein equations

$$\mathcal{R}_{ab} - \frac{1}{2}\mathcal{R}g_{ab} = \kappa T_{ab}, \tag{2.26}$$

contracting the above equation produces $\mathcal{R} - \frac{1}{2}4\mathcal{R} = \kappa T \Rightarrow -\mathcal{R} = \kappa T$ where $T := T^{a}{}_{a}$ and we are considering a four dimensional spacetime. Substituting this result into Eq. (2.26) gives

$$\mathcal{R}_{ab} = \kappa(T_{ab} - \frac{1}{2}T g_{ab}), \tag{2.27}$$

with the vacuum Einstein equation being empty space

$$T_{ab} = 0 \Longrightarrow T = 0 \Longrightarrow \mathcal{R}_{ab} = 0. \tag{2.28}$$

The vacuum solutions are thus characterised by the vanishing of the Ricci tensor.

2.7 Newtonian Limit

Given that we know Newtonian gravitational theory works very well on the scale of the solar system, it is important that we are able to derive Newtonian gravitational theory as an appropriate limit of the Einstein field equations. It is not immediately apparent that this is the case but if it were possible to show that it can't be done, the Einstein field equation would be immediately disproved by the well-tested observations of motion of the planets in the Solar System. We will now obtain the Newtonian limit in the linearised (weak-field) static case.

Consider motion of matter at slow speeds in a weak static gravitational field. Then firstly, there exist coordinates in which the spacetime metric nearly has the Minkowski form,

$$g_{ab} = \eta_{ab} + h_{ab}, \quad |h_{ab}| \ll 1, \tag{2.29}$$

where η_{ab} has a signature $(-1, 1, 1, 1)$ and we set the speed of light $c = 1$ by the appropriate choice of units for distance. Secondly we assume that matter moves slowly compared with the speed of light in these coordinates,

$$\left|\frac{v^\alpha}{c}\right| = \left|\frac{dx^\alpha}{dt}\right| \ll 1. \tag{2.30}$$

Thirdly, all the metric derivatives are small in this coordinate system, and the gravitational field is considered static

$$|h_{ab,c}| \ll 1, \quad h_{ab,c} = 0, \implies |\Gamma^\alpha_{\ bc}| \ll 1. \tag{2.31}$$

All three these conditions are satisfied to a high degree of accuracy in the solar system.

Now we have to check two separate things. The first being that under these conditions we can express particle motion in terms of a gravitational potential Φ, as we can in Newtonian gravitational theory, in the form

$$\frac{d^2x^\alpha}{dt^2} = -\delta^{\alpha\beta}\Phi_{,\beta}. \tag{2.32}$$

Secondly, we need to show that the gravitational equations give the right form of the potential from a given matter distribution, i.e. that Φ satisfies the Poisson equation

$$\nabla^2\Phi = \Phi_{,\alpha\beta}\delta^{\alpha\beta} = 4\pi G\rho, \tag{2.33}$$

where ρ is the matter density. In both these scenarios we approximate by keeping only the largest terms in each equation and consider them symmetric at all stages.

2.7.1 The equations of motion

Consider a test particle moving in a gravitational field satisfying Eq. (2.29)-(2.31), under gravity and inertia alone. Then it moves on a spacetime geodesic, thus obeying

$$\frac{d^2x^a}{d\tau^2} + \Gamma^a_{\ bc}\frac{dx^b}{d\tau}\frac{dx^c}{d\tau} = 0, \tag{2.34}$$

where τ is proper time along the curve. Separating out the space and time terms in the summation,

$$\frac{d^2x^a}{d\tau^2} + \Gamma^a{}_{00}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} + 2\Gamma^a{}_{0\gamma}\frac{dx^0}{d\tau}\frac{dx^\gamma}{d\tau} + \Gamma^\alpha{}_{\beta\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau} = 0.$$

By the slow-motion assumption, Eq (2.30), we can ignore the third and fourth terms, obtaining

$$\frac{d^2x^a}{d\tau^2} + \Gamma^a{}_{00}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} = 0.$$

Now $\Gamma^a{}_{00}$ is given by Eq. (2.21) to be

$$\Gamma^a{}_{00} = g^{ac}\frac{1}{2}\{g_{c0,0} + g_{0c,0} - g_{00,c}\} = g^{ac}\frac{1}{2}\{-h_{00,c}\} = \eta^{ac}\frac{1}{2}h_{00,c},$$

to the first order in h_{ab}, where we have used Eq. (2.31), Eq. (2.29) and the fact that $g^{ab} = \eta^{ab}$ to the required accuracy. Thus the linearised equation of motion is

$$\frac{d^2x^a}{d\tau^2} - \eta^{ac}\frac{1}{2}h_{00,c}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} = 0. \tag{2.35}$$

Putting $a = 0$ in this equation we find from Eq. (2.31),

$$\frac{d^2x^0}{d\tau^2} = \frac{d^2t}{d\tau^2} = 0 \Longrightarrow \frac{dt}{d\tau} = \gamma = \text{const} \simeq 1 \tag{2.36}$$

where in the last relation we have used the definition of $\gamma = 1/\sqrt{1 - v^2/c^2}$, and the slow-motion condition, Eq. (2.30). This is just the result that, for the slow-motion considered, proper time for the particle is the same as the (almost Minkowski) coordinate time, to high accuracy. Using this result in the $a = \alpha$ component of Eq. (2.35),

$$\frac{d^2x^\alpha}{dt^2} = \delta^{\alpha\gamma}\frac{1}{2}h_{00,\gamma}. \tag{2.37}$$

Comparing this result with Eq. (2.32), we see that we have the same equation of motion as in Newtonian gravitational theory where we identify

$$\Phi = -\frac{1}{2}h_{00} \Longleftrightarrow g_{00} = -1(1 + 2\Phi). \tag{2.38}$$

Noting that the static weak field metric is

$$ds^2 = -(1 + 2\phi)dt^2 + (1 - 2\phi)(dx^2 + dy^2 + dz^2). \tag{2.39}$$

2.7.2 The gravitational equations

To obtain the gravitational field equations (2.33), consider equation (2.27) with $a = 0, b = 0$

$$\mathcal{R}_{00} = \kappa(T_{00} - \frac{1}{2}Tg_{00}). \tag{2.40}$$

To simplify the above equation we use the perfect fluid form, $T^{ab} = (\mu + p)u^a u^b + pg^{ab}$, along with a vanishing pressure, this combined with Eq. (2.36) produces

$$T_{ab} = \rho u_a u_b, \quad u^a \simeq (1, v^\alpha) \Longrightarrow T = T^a_{\ a} = -\rho, \quad u_a = g_{ab}u^b \simeq (-1, v^\alpha),$$

this leads Eq. (2.40) to become

$$\mathcal{R}_{00} \approx \kappa(\rho u_0 u_0 + \frac{1}{2}\rho\eta_{00}) = \frac{1}{2}\kappa\rho. \tag{2.41}$$

The left-hand side comes from Eq. (2.23)

$$\mathcal{R}_{00} = \Gamma^s{}_{00,s} - \Gamma^s{}_{s0,0} + \Gamma^e{}_{00}\Gamma^s{}_{se} - \Gamma^e{}_{s0}\Gamma^s{}_{0e}.$$

Now the second term here is zero, as the solution is static, and the third and fourth are second order, since they are products of first order terms, so we have

$$\mathcal{R}_{00} = \Gamma^0{}_{00,0} + \Gamma^\kappa{}_{00,\kappa}.$$

Using Eq. (2.34) and Eq. (2.31) in this equation,

$$\mathcal{R}_{00} = -(\delta^{\kappa\beta}\frac{1}{2}h_{00,\beta})_{,\kappa}.$$

Putting this in Eq. (2.41) and using $h_{00} = -2\Phi$, from Eq. (2.38),

$$\delta^{\kappa\beta}\Phi_{,\beta\kappa} = \frac{1}{2}\kappa\rho. \tag{2.42}$$

This is the same as Eq. (2.33) provided we make the identification

$$\kappa = 8\pi G. \tag{2.43}$$

Thus we have obtained the Newtonian equation (2.33) from (2.27), as desired, and also identified the constant κ in (2.27) in terms of the Newtonian gravitational constant G. This has been done using units such that the speed of light is 1, in general units dimensional analysis shows that relation (2.43) takes the form $\kappa = 8\pi G/c^4$.

This result has come from the "(00)" portion of the Einstein field equations. What we have not done here is show that in the Newtonian limit, the other nine field equations are necessarily satisfied to the appropriate order when we linearise those equations as well. Here we will simply assume it is true. If this were not so, the Newtonian limit would not work out as required, because we would have to satisfy other equations in addition to the Poisson equation (2.42). The theory would not reduce to one effective gravitational field equation.

The implication of the above is that it has been shown that the Einstein field equations, in static, weak field, and slow-motion limit, lead to the Newtonian equations of gravity and the Newtonian particle motions. This is essential in order to ensure that the Einstein theory be a viable theory of gravity. It is a remarkable feature of the equations that the ten Einstein

equations give the simple Newtonian gravitational theory from an appropriate limit, thereby ensuring that all the evidence for the Newtonian theory is also evidence supporting the Einstein theory.

2.8 Exact Solution

Any spacetime metric can in a sense be regarded as satisfying Einstein's field equations

$$\mathcal{R}_{ab} - \frac{1}{2}\mathcal{R}g_{ab} + \Lambda g_{ab} = 8\pi T_{ab} \tag{2.44}$$

where we have used an arbitrary metric tensor (M, g), this will in turn produce an arbitrary T_{ab}. The matter tensor in this form may be unreasonable and not satisfy normal physical properties.

In defining an exact solution to the Einstein field equations we are looking for a solution in which the metric tensor (M, g), produces an energy-momentum solution of a specific form of matter which adheres to local causality and energy conditions.

High degrees of symmetry in the spacetime are easier to analyse and make it simpler to solve the complex nature of the field equations.The exact solutions produce only simple matter content. This is an idealised situation since one can have a region of spacetime containing varying forms of matter but these simplified solutions help produce a qualitative understanding of the features of the spacetime. The examples we explore will cover the global properties. The discussion of the different exact solutions in the next section uses [HE73] as a guideline for the structure.

Before moving on to the exact solutions this is a good point to review and put into context the differences between Newtonian gravity and general relativity. This is dealt with by laying out the comparison in a tabular form, Table 2.8.

Table 2.1: A comparison of Newtonian Gravity and General Relativity. Taken from [Har03].

	Newtonian Gravity	**General Relativity**
What mass does	Produces a field Φ causing a force on other masses $F = -m\boldsymbol{\nabla}\phi$	Curves spacetime $ds^2 = g_{ab}(x)dx^a dx^b$
Motion of a particle	Newton's law of motion $\frac{d^2x^i}{dt^2} + \delta^i_j \frac{\partial \phi}{\partial x^j} = 0$	Geodesic equation $\frac{d^2x^a}{d\tau^2} + \Gamma^a_{bg}\frac{dx^b}{d\tau}\frac{dx^g}{d\tau} = 0$
Field equation	Newtonian field equation $\boldsymbol{\nabla}^2\phi = 4\pi\nu$	The Einstein equation $\mathcal{R}_{ab} - \frac{1}{2}\mathcal{R}g_{ab} = 8\pi T_{ab}$

2.9 Minkowski Solution

The simplest empty spacetime of General Relativity is Minkowski spacetime (M, g) and is equivalent to the spacetime of Special Relativity. To achieve this in terms of the Einstein equations we set $T_{ab} = 0$. This spacetime is of course only applicable to idealised situations.

It is a manifold $\mathbb{R}^4$ with flat Lorentz metric $\eta = \text{Diag}(-1, 1, 1, 1)$,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{2.45}$$

If one uses spherical polar coordinates (t, r, θ, ϕ) where $x^4 = t,\ x^3 = r\cos\theta,\ x^2 = r\sin\theta\cos\phi,\ x^1 = r\sin\theta\sin\phi$, the metric takes the form

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.46}$$

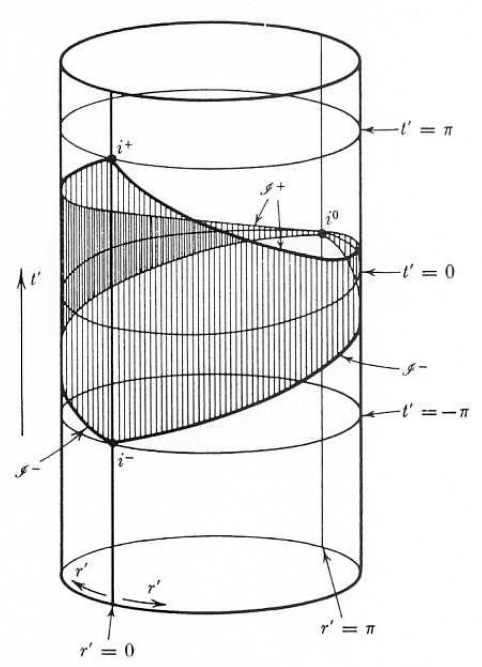


Figure 2.5: The Einstein static universe represented by an embedded cylinder; the coordinates θ, ϕ have been suppressed. Each point represents one half of a two-sphere of area $4\pi\sin^2 r'$. The shaded region is conformal to the whole of Minkowski spacetime; its boundary (part of the null cones of i^+, i^0 and i^-) may be regarded as the conformal infinity of Minkowski spacetime. This diagram is taken from [HE73].

At this point it is of value to digress slightly. Penrose gives a representation that better enables us to understand the structure of infinity in Minkowski spacetime. This allows us to visualise infinite values that are transformed to finite values. We thus have a representation of infinity. In general a Penrose diagram is used to describe the causal global structure of spacetime. The major feature of these diagrams is to put the infinity at a finite position while simultaneously preserving the null geodesics. By doing so one can study the global structure and the infinity.

Only radial and time directions are represented while angular directions are suppressed. A Penrose diagram is not meant to accurately portray distances, only causal structure. Light

rays travel at 45^o angles and the metric on the Penrose diagram is conformally equivalent to the actual metric so $ds^2 = 0$ is preserved. Each point in a Penrose diagram represents a two-sphere, except for i^+, i^0 and i^- (defined below) which are each single points along with the line $r = 0$ where the polar coordinates are singular.

In order to construct a Penrose diagram we first select an alternate coordinate system by choosing advanced and retarded coordinates v, w defined by $v = t + r, w = t - r (\Rightarrow v \leq w)$. This means that the metric becomes

$$ds^2 = -dv\ dw + \frac{1}{4}(v-w)^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{2.47}$$

where $-\infty < v < \infty, -\infty < w < \infty$. We now define p, q by $\tan p = v, \tan q = w$ where $-\frac{1}{2}\pi < p < \frac{1}{2}, -\frac{1}{2}\pi < q < \frac{1}{2}$(and $p \leq q$). This produces a metric of (M, η) that takes the form

$$d\bar{s}^2 = -4dp\ dq + \sin^2(p-q)(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.48}$$

This metric can be reduced to a more usual form by defining

$$t' = p + q, \quad r' = p - q, \quad \text{where} - \pi < t' + r' < \pi, -\pi < t' - r' < \pi, r' \geq 0, \tag{2.49}$$

this is then,

$$d\bar{s}^2 = -(dt')^2 + (dr')^2 + \sin^2 r'(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.50}$$

Thus the whole Minkowski spacetime is given by the region in Eq. (2.49) by the metric

$$ds^2 = \frac{1}{4}\sec^2\left(\frac{1}{2}(t'+r')\right)\sec^2\left(\frac{1}{2}(t'-r')\right)d\bar{s}^2 \tag{2.51}$$

where $d\bar{s}^2$ is determined by Eq. (2.50); the coordinates t, r of Eq. (2.46) are related to t', r' by

$$2t = \tan\left(\frac{1}{2}(t'+r')\right) + \tan\left(\frac{1}{2}(t'-r')\right), \quad 2r = \tan\left(\frac{1}{2}(t'+r')\right) + \tan\left(\frac{1}{2}(t'-r')\right). \tag{2.52}$$

Now the metric Eq. (2.21) is locally identical to that of the Einstein static universe which is a completely homogeneous spacetime. One can analytically extend Eq. (2.50) to the whole of the Einstein static universe, that is one can extend the coordinates to cover the manifold $\mathbb{R}^1 \times S^3$ where $-\infty < t' < \infty$ and r', θ, ϕ are regarded as coordinates on S^3 (with coordinate singularities at $r' = 0, r' = \pi$ and $\theta = 0, \theta = \pi$). On suppressing two dimensions, one can represent the Einstein static universe as the cylinder $x^2 + y^2 = 1$ embedded in a three dimensional Minkowski space with metric $ds^2 = -dt^2 + dx^2 + dy^2$. The full Einstein static universe can be embedded as the cylinder $x^2 + y^2 + z^2 + w^2 = 1$ in a five dimensional Euclidean space with metric $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 + dw^2$, [HE73].

One therefore has the situation where the whole of the Minkowski spacetime is conformal to the region given in Eq. (2.49) of the Einstein static universe, this is illustrated as the shaded area in Figure 2.5. The boundary of this region may therefore be thought of as representing the conformal structure of infinity of Minkowski spacetime. It consists of the null surfaces $p = \frac{1}{2}\pi$ (labelled $\mathcal{I}^+$) and $q = -\frac{1}{2}$ (labelled $\mathcal{I}^-$) together with points $p = \frac{1}{2}\pi, q = -\frac{1}{2}\pi$ (labelled

i^+),$p = \frac{1}{2}\pi, q = -\frac{1}{2}\pi$ (labelled i^0) and $p = -\frac{1}{2}\pi, q = -\frac{1}{2}\pi$ (labelled i^-). Any future directed timelike geodesic in Minkowski space approaches $i^+(i^-)$ for indefinitely large positive (negative) values of its affine parameter, so one can regard any timelike geodesic as originating at i^- and finishing at i^-, as in Figure 2.6 (i). Similarly one can regard null geodesics as originating at $\mathcal{I}^-$ and ending at $\mathcal{I}^+$, while spacelike geodesics both originate and end at i^0. Thus one may regard i^+ and i^- as representing future and past timelike infinity, and i^0 as representing spacelike infinity.

One can also represent the conformal structure of infinity by drawing a diagram of the (t', r') plane, see Figure 2.6 (ii). Each point of this diagram represents a sphere S^2, and radial null geodesics are represented by straight lines at $\pm 45^o$. In fact, the structure of infinity in any spherically symmetric spacetime can be represented by a Penrose diagram. On such diagrams we shall represent infinity by single lines, the origin of polar coordinates by dotted lines, and irremovable singularities of the metric by double lines. One can obtain spaces locally identical to (M, η) but with differing large scale topological properties by identifying points in M which are equivalent under a discrete isometry without a fixed point. An example of this would be identifying the points (x^1, x^2, x^3, x^4) with the point $(x^1, x^2, x^3, x^4 + c)$, where c is a constant. This changes the topological structure from $\mathbb{R}^4$ to $\mathbb{R}^3 \times S^1$, and introduces closed timelike lines into the spacetime. (M, η) is clearly then the universal covering space for all such derived spaces.

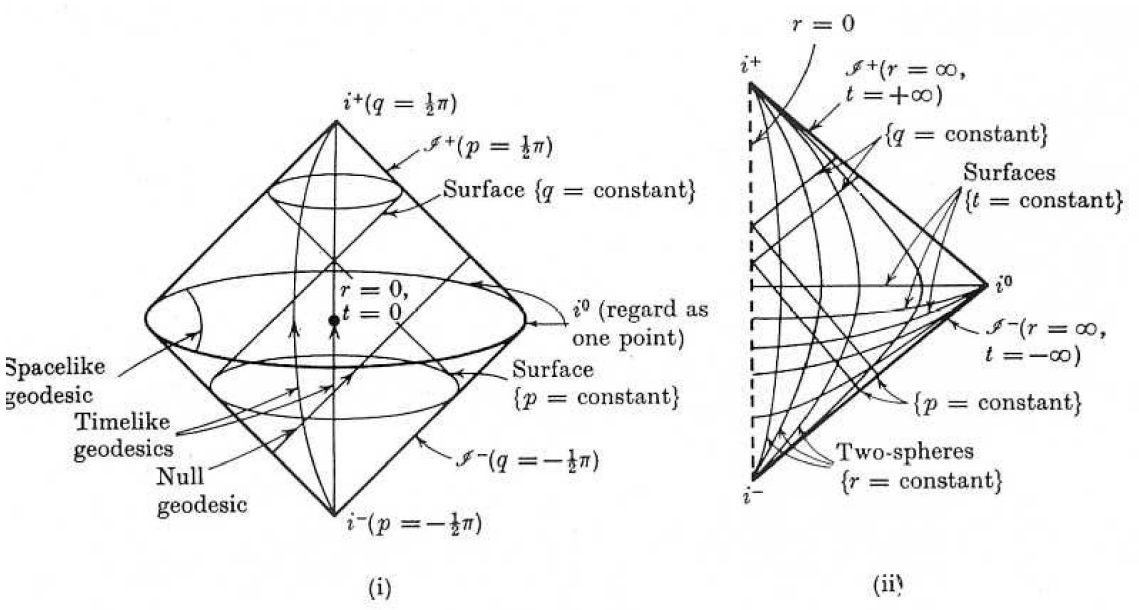


Figure 2.6: (i) The shaded region of Figure 2.5, with only one coordinate suppressed, representing Minkowski spacetime and its conformal infinity. (ii) The Penrose diagram of Minkowski spacetime; each point represents a two-sphere, except for i^+, i^0 and i^-, each of which is a single point, and points on the line $r = 0$ (where the polar coordinates are singular). This diagram is taken from [HE73].

2.10 De Sitter Solution

This is a spacetime that has a constant curvature and is characterised by being homogeneous. The Einstein tensor is

$$\mathcal{R}_{ab} - \frac{1}{2}\mathcal{R}g_{ab} = -\frac{1}{4}\mathcal{R}g_{ab}. \tag{2.53}$$

This spacetime thus represents a solution for the Einstein field equations with $\Lambda = \frac{1}{4}\mathcal{R}$ where $\mathcal{R} > 0$ or put another way a solution with a positive cosmological constant but with no matter content. Other spaces with a constant curvature are Minkowski $\mathcal{R} = 0$ and anti de Sitter $\mathcal{R} < 0$.

These features make de Sitter a natural candidate for an expanding universe. The question then arises, does it 'fit' our universe? Since we have observed $\Lambda \neq 0$, with $\Lambda \approx \rho_{\text{darkmatter}}$, and Λ is constant while ρ is decreasing with the expansion of the universe, soon ρ will be negligible and the universe will be described by approximately de Sitter spacetime. Although de Sitter is used in the context of an expanding universe it should be noted that there is no Big Bang in a de Sitter universe. Expansion is self-similar from an infinite time in the past to an infinite time in the future.

De Sitter spacetime has a topology of $\mathbb{R}^1 \times S^3$ and can be best visualised as a hyperboloid

$$-v^2 + w^2 + x^2 + y^2 + z^2 = \alpha^2 \tag{2.54}$$

in flat five-dimensional space, $\mathbb{R}^5$, with metric

$$-dv^2 + dw^2 + dx^2 + dy^2 + dz^2 = ds^2. \tag{2.55}$$

This is illustrated in Figure 2.7 (i).

Introducing the following coordinates

$$\begin{aligned}
v &= \alpha \sinh(\alpha^{-1}t), \\
w &= \alpha \cosh(\alpha^{-1}t), \\
x &= \alpha \cosh(\alpha^{-1}t) \sin\chi \cos\theta, \\
y &= \alpha \cosh(\alpha^{-1}t) \sin\chi \sin\theta \cos\phi, \\
z &= \alpha \cosh(\alpha^{-1}t) \sin\chi \sin\theta \sin\phi,
\end{aligned}$$

produces the following form for the metric,

$$ds^2 = -dt^2 + \alpha^2 \cosh^2(\alpha^{-1}t)\{d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\}. \tag{2.56}$$

Another set of coordinates that are common can also be introduced on the hyperboloid

$$\begin{aligned}
\hat{t} &= \alpha \log\frac{w+v}{\alpha}, \\
\hat{x} &= \frac{\alpha x}{w+v}, \\
\hat{y} &= \frac{\alpha y}{w+v}, \\
\hat{z} &= \frac{\alpha z}{w+v}.
\end{aligned}$$

The metric then takes the form

$$ds^2 = -d\hat{t}^2 + \exp(2\alpha^{-1}\hat{t})(d\hat{x}^2 + d\hat{y}^2 + d\hat{z}^2). \tag{2.57}$$

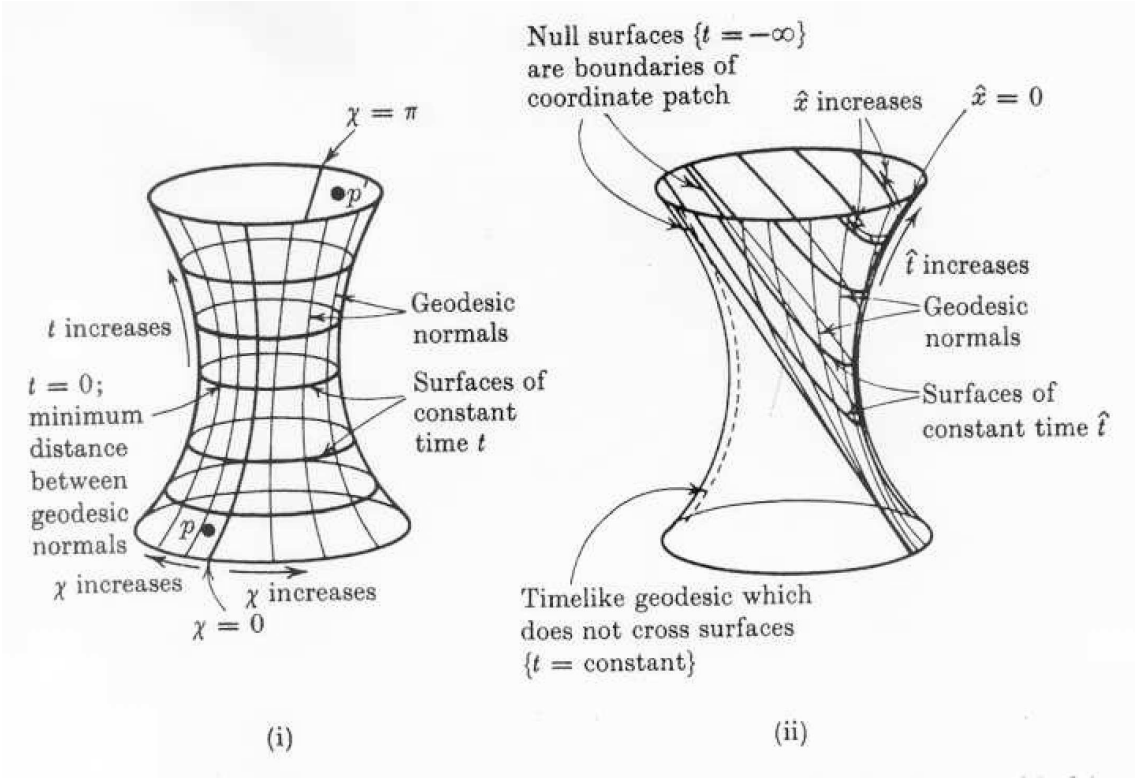


Figure 2.7: De Sitter spacetime represented by a hyperboloid embedded in a five-dimensional flat space (two dimensions are suppressed in the figure). (i) Coordinates (t, χ, θ, ϕ) cover the whole hyperboloid; the sections $\{t = \text{constant}\}$ are surfaces of curvature $k = +1$. (ii) Coordinates $(\hat{t}, \hat{x}, \hat{y}, \hat{z})$ cover half the hyperboloid; the surfaces $\{\hat{t} = \text{constant}\}$are flat three-spaces, there geodesic normals diverging from a point in the infinite past. This diagram is taken from [HE73].

These coordinates only cover the region where $w + v \leq 0$ of the hyperboloid. This can be more clearly seen in Figure 2.7 (ii).

To develop a Penrose diagram of the de Sitter space we need to define a new time coordinates, t',

$$t' = 2 \text{ arctan } (\exp \alpha^{-1} t) - \frac{1}{2}\pi, \tag{2.58}$$

in order to study infinity, where

$$-\frac{1}{2}\pi < t' < \frac{1}{2}\pi.$$

Then

$$ds^2 = \alpha^2 \cosh^2(\alpha^{-1}) d\bar{s}^2, \tag{2.59}$$

where $d\bar{s}^2$ is given by Eq. (2.50) when we identify $r' = \chi$. We can see a conformal relationship, Figure 2.8 (i), to the Einstein static universe given by Eq. (2.10). This leads us the Penrose diagram in Figure 2.8 (ii) and Figure 2.8 (iii). A distinction that can be drawn between de Sitter and Minkowski spaces by comparing the Penrose diagrams is that de Sitter space has both, future and past directed, spacelike infinities for timelike and null lines, where as Minkowski space has no spacelike infinity. This has implications that are discussed below.

In Figure 2.9 (i) we consider a family of timelike geodesics in de Sitter space. The geodesics

originating at the spacelike infinity, $\mathcal{I}^-$ and on the spacelike infinity, $\mathcal{I}^+$. The point p is an event on the worldline of O. The past null cone of p represents the past observable events by O at the point p. The other worldlines that intersect this null cone are particles that are visible to O. From the diagram we can see that there may exist worldlines that fall outside the past null cone of p and thus are not visible to O. This division of unseen and seen particles represents the particle horizon for the observer O and the event p.

By contrast in Figure 2.9 (ii), considering Minkowski space, all the other particles are visible at any point on O's world line, if they move on timelike geodesics. If we thus only consider families of geodesic observers then the existence of the particle horizon is a direct consequence of having a past null infinity being spacelike.

On the limit of O's world line at $\mathcal{I}^+$ the past null cone forms a boundary between events that are visible by O and those that will never be observable. This is called the *future event horizon* of the world line. In Minkowski spacetime, however, the limiting null cone of any geodesic observer includes the whole spacetime. There are thus no events that a geodesic observer will never be able to see. The only case where a horizon can be constructed in Minkowski space is if we consider an observer moving with a uniform acceleration in which case a future event horizon may exist, Figure 2.10 (ii). One may think of a future event horizon being a consequence of having $\mathcal{I}^+$ being spacelike.

The future null cone on O's world line is the boundary of the set of events in spacetime which O can influence. The maximal set of events that O can influence at any time in the spacetime is chosen when we take the future null cone (creation light cone) at the point of the past infinity $\mathcal{I}^-$ of the worldline O. The region intersecting the future event horizon and the past event horizon, is often referred to as the causal diamond, for apparent reasons.

A local equivalent space to de Sitter can be constructed by identifying points. The simplest identification is noting antipodal points p and p', shown in Figure 2.7, on the hyperboloid. The resulting space is not time orientable, if time increases in the direction of the arrow at p', but one cannot continuously extend this identification of future and past half null cones over the whole hyperboloid. A particular attribute of this is that any two points in the resulting space can be joined to any other points by geodesics if and only if it is not time orientable.

2.11 Anti-de Sitter Solution

Anti-de Sitter space has the opposite constant curvature $\mathcal{R} < 0$ to de Sitter space $\mathcal{R} > 0$. It has a topology $S^1 \times \mathbb{R}^3$ and can be represented as a hyperboloid

$$-u^2 - v^2 + x^2 + y^2 + z^2 = 1$$

in the flat five dimensional space $\mathbb{R}^5$ with the metric

$$ds^2 = -(du)^2 - (dv)^2 + (dx)^2 + (dy)^2 + (dz)^2.$$

This space has a universal covering space which is given by $\mathbb{R}^4$ and does not contain any closed timelike lines. It is this universal covering space which shall be referred to as anti-de Sitter space

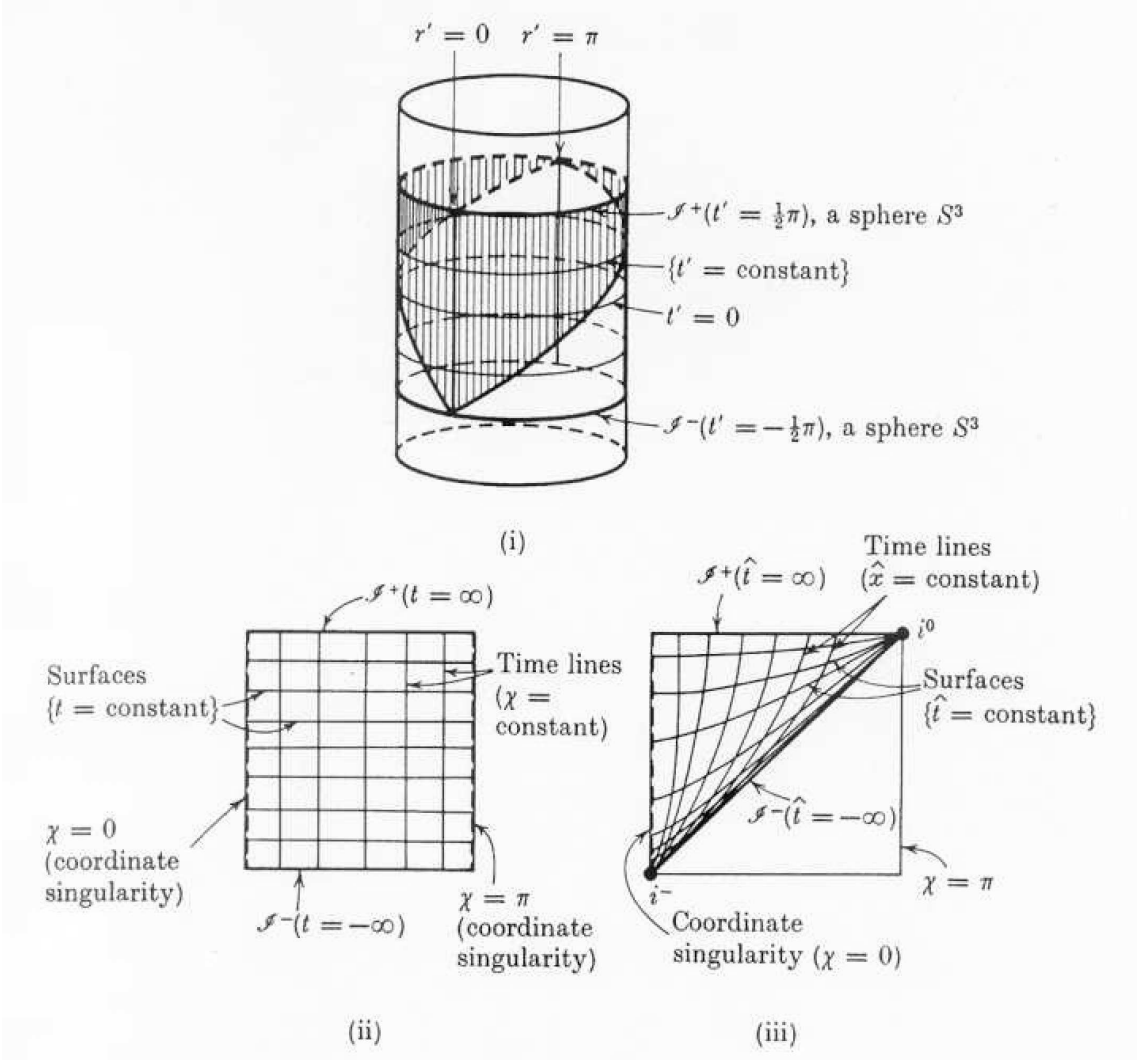


Figure 2.8: (i) De Sitter spacetime is conformal to the region $-\frac{1}{2} < t' < \frac{1}{2}\pi$ of the Einstein static universe. The steady state universe is conformal to the shaded region. (ii) The Penrose diagram of de Sitter spacetime. (iii) The Penrose diagram of the steady state universe. In (ii), (iii) each point represents a two sphere of area $2\pi \sin^2 \chi$; null lines are at 45^o. $\chi = 0$ and $\chi = \pi$ are identified. This diagram is taken from [HE73].

in future. The following metric can represent this case,

$$ds^2 = -dt^2 + \cos^2 t d\chi^2 + \sinh^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2). \tag{2.60}$$

There appears to be a singularity at $t = \pm\frac{1}{2}\pi$ while Eq. (2.60) only covers a portion of the space. This is the diamond shaped region in Figure 2.11 (i). The entire space can be covered successfully, shown in Figure 2.11 (i), by the coordinates t', r, θ, ϕ,

$$ds^2 = -\cosh^2 r dt'^2 + dr^2 + \sinh^2 r (d\theta^2 + \sin^2 \theta d\phi^2).$$

This allows us to cover the space with the surfaces where $t' =$ constant.

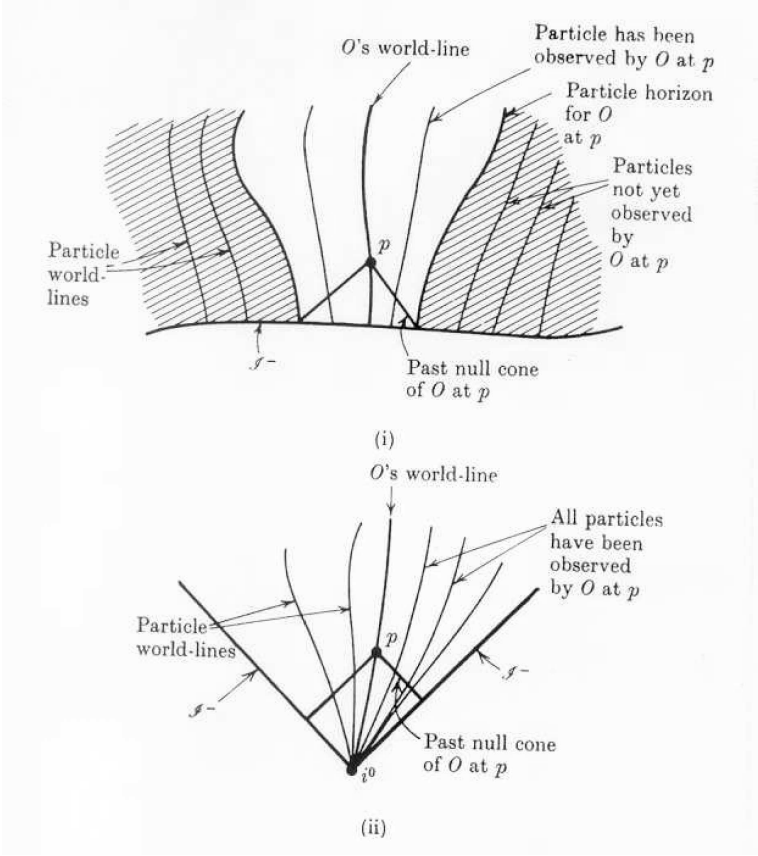


Figure 2.9: (i)The particle horizon defined by a congruence of geodesics curves when the past null infinity I^- is spacelike. (ii) Lack of such a horizon if I^- is null. This diagram is taken from [HE73].

The Penrose diagram is shown in Figure 2.11 (ii). Here we have defined

$$r' = 2\arctan(\exp r) - \frac{1}{2}\pi, \quad 0 \leq r' < \frac{1}{2}\pi,$$

such that we can consider infinity. We can now rewrite the metric as

$$ds^2 = \cosh^2 d\bar{s}^2 = -(dt')^2 + (dr')^2 + \sin^2 r'(d\theta^2 + \sin^2\theta d\phi^2).$$

Then one finds $ds^2 = -\cosh^2 r d\bar{s}^2$, where $d\bar{s}$ is the same as given in the section on the Minkowski solution, Eq. (2.50). This coordinate solution is thus conformal to the region $0 \leq t' \leq \frac{1}{2}\pi$ of the Einstein static cylinder. One of the distinguishing points about this space is that null and spacelike infinity can be thought of as timelike surfaces in this case. The topology in this instance is $\mathbb{R}^1 \times S^2$. A complication that arises is that there is no way of finding a conformal mapping representing timelike infinity as finite without reducing the Einstein static universe to a point. This is overcome by making the disjoint points i^+ and i^- represent the timelike infinity. The geodesics orthogonal to the surfaces $t =$ constant are the lines $\{\chi, \theta, \phi$ constant$\}$ and they all converge to the point q in the future or p in the past.

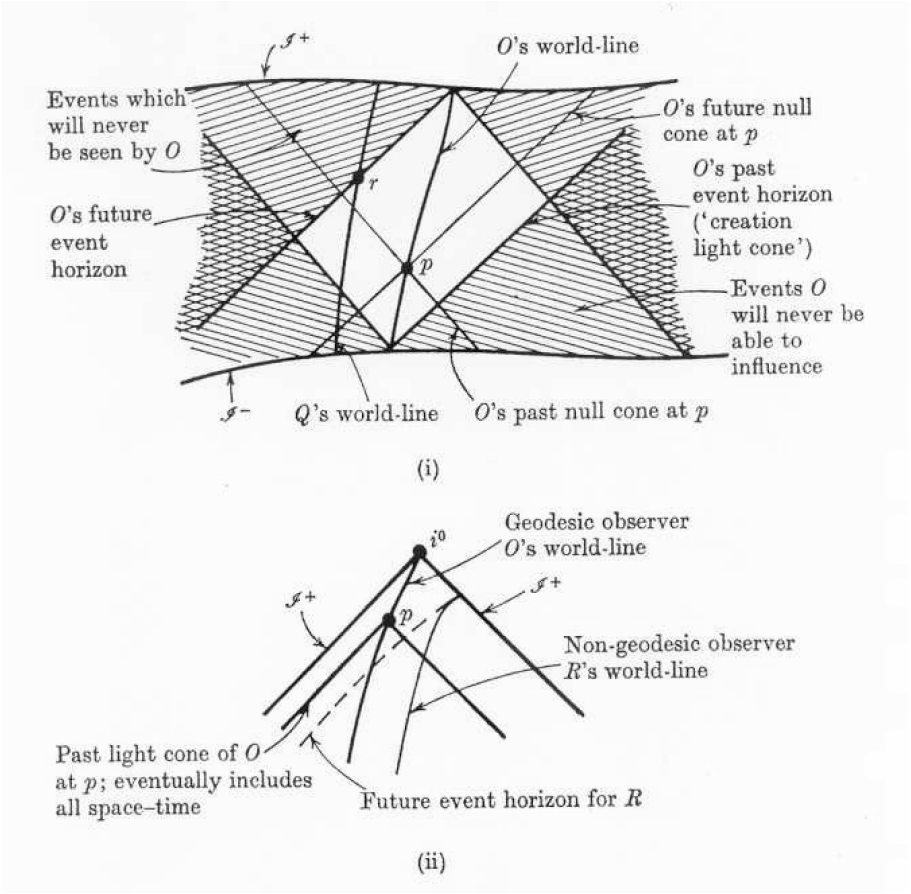


Figure 2.10: (i) The future event horizon for a particle O which exists when the future infinity I^+ is spacelike; also the past event horizon which exists when the past infinity I^- is spacelike. (ii) If the future infinity consists of a null I^+ and i^0, there is no future event horizon for a geodesic observer O. However an accelerating observer R may have a future event horizon. This diagram is taken from [HE73].

2.12 Robertson Walker Solutions

The Robertson Walker solution is a model which is able to represent the physical universe. It requires that the distribution is homogeneous (the same all around us) and isotropic (the same in all directions). Observation supports the two previous conditions in our universe, it is also a solution with a constant curvature. This is thus a good approximation to the large scale geometry of spacetime in the region we can observe. Minkowski space, de Sitter and Anti-de Sitter are all special cases of the general Friedman Robertson Walker spaces and will be touched on again at a later stage.

The Robertson Walker solution is referred to as the Friedman Robertson Walker (FRW) solution when the scale factor obeys the Einstein equation, [Har03]. One of the attributes of the Friedman Robertson Walker model is that the distribution of the galaxies are smoothed into a cosmological fluid[1]. A single galaxy is thought of as a particle in a fluid located by three coordinates at any time. These coordinates are flow lines given by the curves (χ, θ, ϕ) and are therefore comoving coordinates. The function $S(t)$ represents the separation of neighbouring

[1]This requires that the density, μ, and pressure, p are functions of time only.

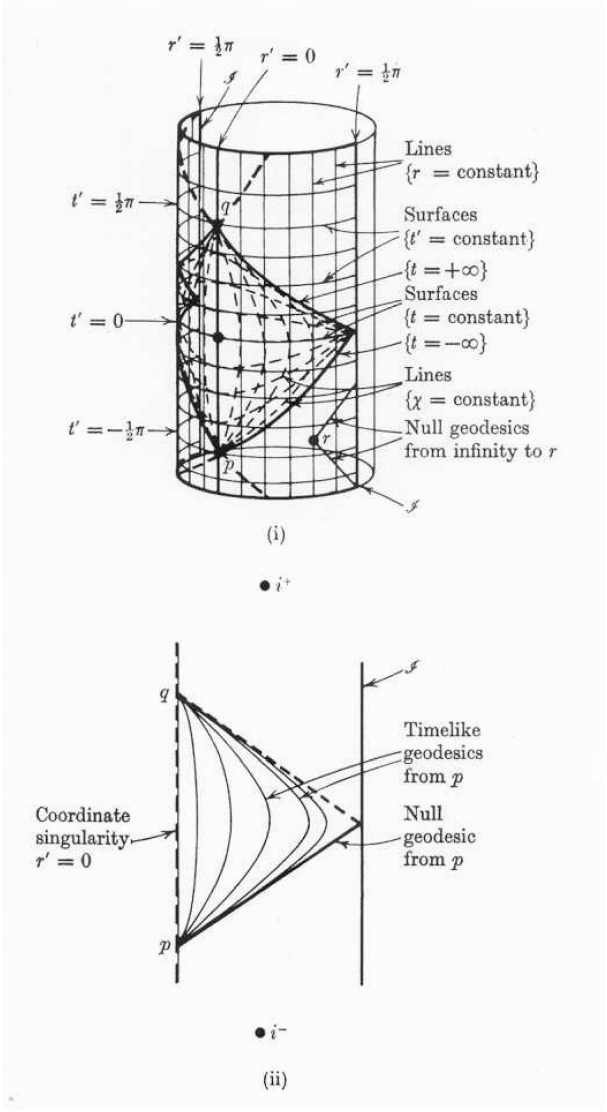


Figure 2.11: (i) Universal anti-de Sitter space is conformal to one half of the Einstein static universe. While coordinates (t', r, θ, ϕ) cover the whole space, coordinates (t, χ, θ, ϕ) cover only one diamond shaped region as shown. The geodesics diverge out into similar diamond shaped regions. (ii) The Penrose diagram of universal anti-de Sitter space. Infinity consists of the timelike surface I and the disjoint points i^+, i^-. The projection of some timelike and null geodesics is shown. This diagram is taken from [HE73].

galaxies or flow lines.

Coordinates can be chosen to express the metric as

$$ds^2 = -dt^2 + S^2(t)(d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)),$$

where the last term, $d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)$, is a metric for a three-space of constant curvature and is independent of time. The geometry of this term can have three possibilities depending on whether the curvature is positive, negative or zero. Homogeneity requires that the spatial curvature be the same at each point of these geometries. Consider K to be the normalised

curvature constant then the following options exist,

$$f(\chi) = \begin{cases} \sin\chi & \text{if} \quad K = +1, \\ \chi & \text{if} \quad K = 0, \\ \sinh\chi & \text{if} \quad K = -1. \end{cases}$$

These varying solutions given by K are referred to as flat, when the curvature is zero, closed when the curvature is negative and open when the curvature is positive. Figure 2.12 shows how the open, flat and closed geometries can be conceptualised for the FRW metric.

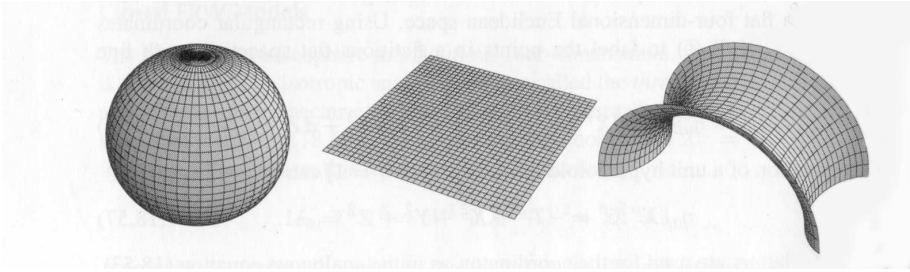


Figure 2.12: All three diagrams shown here are illustrations of homogeneous and isotropic geometries for the space of a FRW metric. These are embedded diagrams, from left to right, of a constant positive curvature (sphere) or $K = +1$, a flat curvature or $K = 0$ (plane) and constant negative curvature (saddle) or $K = -1$. This diagram is taken from [Har03].

The Friedman equation can be derived from the field equation where a pressureless gas is assumed with Λ taken to be zero,

$$\dot{S}^2 - \frac{M}{S} = -K \equiv \frac{E}{M}. \tag{2.61}$$

This is an equation for energy conservation and a comoving volume of matter where the E represents the sum of the kinetic and potential energies. Consider differing values for K and therefore differing energies, this produces varying results for S, which is illustrated in Figure 2.13. We can see that for $K = -1$ or $K = 0$ the scale factor S increases indefinitely while $K = +1$ produces an S that increases to a maximum and then decreases to zero again.

The solutions for the Friedman equation are relatively straightforward when one uses a rescaled time parameter $\tau(t)$ which is defined by

$$\frac{d\tau}{dt} = S^{-1}(t). \tag{2.62}$$

The solutions are listed in [HE73] and the different scenarios of Λ being negative, positive or equal to zero are also discussed.

- Λ negative : The solution expands from an initial singularity (big bang) to a maximum and then recollapses to a singularity again (big crunch).
- Λ positive : There are two possibilities depending on the value of K,

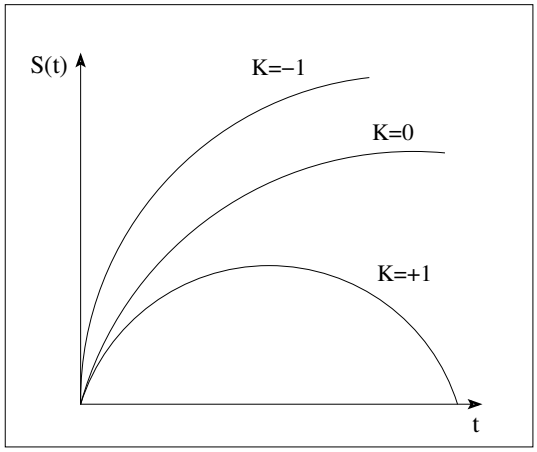


Figure 2.13: A graph of the scale factor, $S(t)$, versus time, t. For $K = -1$ and $K = 0$ it increases indefinitely while for $K = +1$ it increases to a maximum and then decreases to zero again. Based on a diagram from [Har03].

- $K = 0$ or $K = -1$: In this instance the solution expands indefinitely while asymptotically approaching the steady state.
- $K = +1$: There are several solutions depending on the value of Λ relative to $\Lambda_{\text{crit}} = (\frac{-E}{3m^3})/(3M)^2$ if $p = 0$,
  - $\Lambda > \Lambda_{\text{crit}}$: The solution will be similar to that where Λ is negative, i.e. it will expand from an initial singularity and asymptotically approach the static solution.
  - $\Lambda = \Lambda_{\text{crit}}$: This is a static solution, Einstein static universe.
  - $\Lambda < \Lambda_{\text{crit}}$: There are two alternatives depending on the value of the scale factor S. If it is relatively small it will produce a big crunch solution as discussed earlier. On the other hand a large scale factor will result in a solution which collapses from an infinite radius in the past to a minimum radius and then re-expands forever without becoming singular (the bounce).

The conformal mappings of the Robertson Walker spaces into the Einstein static space, can be found by using the coordinate τ as a time coordinate, Eq. (2.62), we arrive at

$$ds^2 = S^2(\tau) - d\tau^2 + d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.63}$$

We now examine each of the cases for K.

- $K = +1$: This solution is already conformal to the Einstein static space, where $\tau = t'$ and $\chi = r'$ as in the notation of Eq.(2.62). Setting $\Lambda = p = 0$ we arrive at a space that is mapped into the Einstein static space where τ lies between $0 < \tau < \pi$. This mapped space thus forms a band, between $t' = 0$ and $t' = \pi$ in the illustration of the Einstein static universe, Figure 2.14 (i).
- $K = 0$: This instance is conformal to flat space we thus expect that it would map into the same region as the Minkowski space mapped into the Einstein static universe. When

$\Lambda = 0$ this solution maps into the upper region, $t' > 0$, only of the Minkowski diamond illustrated in Figure 2.5 and is shown in Figure 2.14 (i).

- $K = -1$: Once we define

$$t' = \arctan(\tanh\frac{1}{2}(\tau + \chi)) + \arctan(\tanh\frac{1}{2}(\tau - \chi)),$$
$$r' = \arctan(\tanh\frac{1}{2}(\tau + \chi)) - \arctan(\tanh\frac{1}{2}(\tau - \chi)),$$

we arrive at a region $\frac{1}{2}\pi \geq t' + r' \geq -\frac{1}{2}\pi$ and $\frac{1}{2}\pi \geq t' - r' \geq -\frac{1}{2}\pi$ that is conformal to a portion of the Einstein static space. When $\Lambda = 0$ we cover only the region where $0 \geq t' \geq \frac{1}{2}\pi$. This is depicted in Figure 2.14 (i).

The Penrose diagram for $K = +1$ where $p = \Lambda = 0$, Figure 2.14 (ii), has a singularity at $t = 0$ which is given by a spacelike surface and this in turn implies that there are particle horizons. The future boundary is also spacelike in this instance which corresponds to the existence of event horizons for the fundamental observer (Both particle and event horizons where discussed in Section 2.10 and illustrated in Figure 2.9 and 2.10). While the case for $K = 0$ or $K = -1$ and $p = \Lambda = 0$, Figure 2.14 (ii), also has a spacelike boundary at $t = 0$, but has a null future infinity, unlike the $K = +1$ case, where event horizons exist for the fundamental observer.

2.13 Schwarzschild Solution

Homogeneous solutions are good for large scale distribution of matter in the universe but are inadequate for describing the local geometry of, for example, the spacetime of the solar system. A solution that gives you a good approximation of this kind of geometry is the Schwarzschild solution. It gives a description of an empty metric that is spherically symmetric near a massive spherically symmetric body, for example a star. Most experiments done testing the difference between general relativity and Newtonian theory have been based on predictions by this solution. Figure 2.15 gives a good introduction to the geometry of a Schwarzschild metric and it can be seen how the curvature is concentrated around a large mass (or a black hole in the case of the diagram) causing a significant distortion at the centre while it asymptotically approaches flat space further away.

The line element summarising the Schwarzschild geometry is given by,

$$ds^2 = -\left(1 - \frac{2m}{r}\right)dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2 d\phi^2), \tag{2.64}$$

where $r > 2m$. There are a few immediate properties that can be deduced from the above metric. One of the most fundamental characteristics of this metric is that it is spherically symmetric and time independent. Examining Eq. (2.64) highlights that there are two obvious singularities at $r = 0$ and $r = 2m$. The latter is called the Schwarzschild radius and in the case of a static star, the surface would always be outside this radius, $r > 2m$. The radius, $r \leq 2m$, only becomes a significant consideration when looking at black holes.

As an interesting illustration of how applicable the above solution is we can consider a small

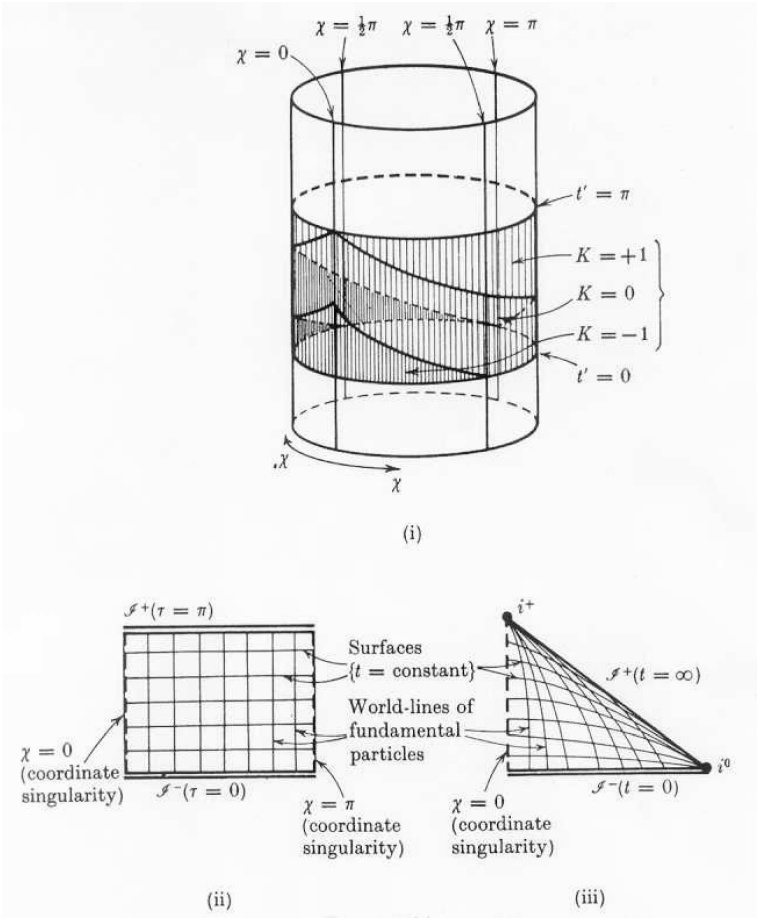


Figure 2.14: (i) The Robertson-Walker spaces $(\rho = \Lambda = 0)$ **are conformal to the regions of static universe shown, in the three cases** $K = +1, 0$ **and** -1**. (ii) Penrose diagram of a Robertson-Walker space with** $K = +1$ **and** $\rho = \Lambda = 0$**. (iii) Penrose diagram of a Robertson-Walker space with** $K = 0$ **or** -1 **and** $\rho = \Lambda = 0$**. This diagram is taken from [HE73].**

value for m/r, then the coefficient of dr^2 in Eq. (2.64) can be expanded to give,

$$ds^2 \approx -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 + \frac{2m}{r}\right) dr^2 + r^2(d\theta^2 + \sin^2 d\phi^2).$$

This corresponds to the static weak field metric, Eq. (2.39), in spherical coordinates with a Newtonian gravitational potential Φ given by,

$$\Phi = -\frac{m}{r}.$$

With a large orbit and a relatively small mass we then arrive at a Newtonian limit where the mass becomes the source of the curvature showing the accuracy of the Schwarzschild solution in a well understood classical setting.

A black hole is a small highly dense concentration of matter that is within the Schwarzschild radius, $r < 2m$, and adheres to the metric Eq. (2.64). The formation of a black hole is caused by the collapse of a star that has burnt a large portion of its nuclear fuel and no longer has enough

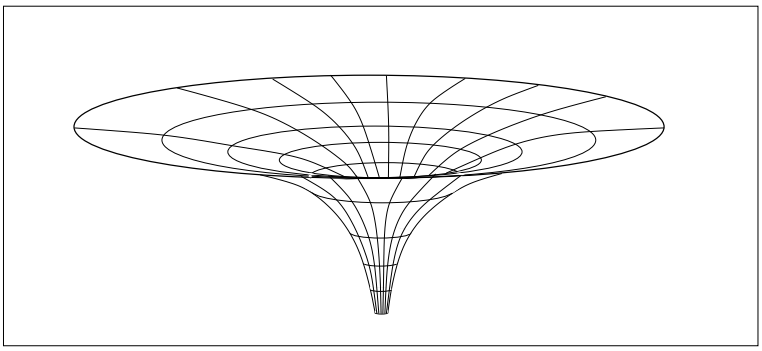


Figure 2.15: This diagram is an embedded depiction of a black hole where one of the spatial dimensions is suppressed. This shows the curvature of the geometry and one can see how it becomes very steep towards the centre of the diagram, corresponding to a strong gravitational field, where the black hole's singularity exists and asymptotically flat further away from the centre. This diagram makes it easier to imagine how a trajectory of a particle would travel on this type of geometry and ultimately could be caught if it was to pass too close to the very steep throat of black hole's geometry. Based on diagram from [Har03].

energy to balance the loss of the stars energy to radiation with its internal gravitational forces resulting in it collapsing on itself until it eventually decreases to the Schwarzschild radius. The formation of a black hole is is depicted in Figure 2.16 along with some of the physical features of a black hole. In the case of a dying star that collapses in on itself there is also a possibility that it reaches a stable state due to non-thermal source of pressure that balance the gravitational forces. This will not result in a black hole and therefore is not of interest to us.

As we are now examining more of the "internal" structure of the Schwarzschild geometry we will have to deal with the singularity at $r = 2m$. Firstly it is important to note that this singularity is in fact a coordinate singularity and although it does have physical significance it can be removed by a wise choice of coordinates and an analytic continuation . We begin by defining,

$$r \equiv \int \frac{dr}{1 - 2m/r} = r + 2m \log(r - 2m),$$

and then

$$v \equiv t + r,$$

which is an advanced null coordinate while,

$$w \equiv t - r,$$

is a retarded null coordinate. This then forms the basis for the Eddington-Finkelstein form of the metric by using the coordinates (v, r, θ, ϕ) which produces,

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dv^2 + 2dvdr + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.65}$$

This is exactly equivalent to the previous metric given in Eq. (2.64) but applies across the range $0 < r < \infty$ and not just where $r > 2m$.

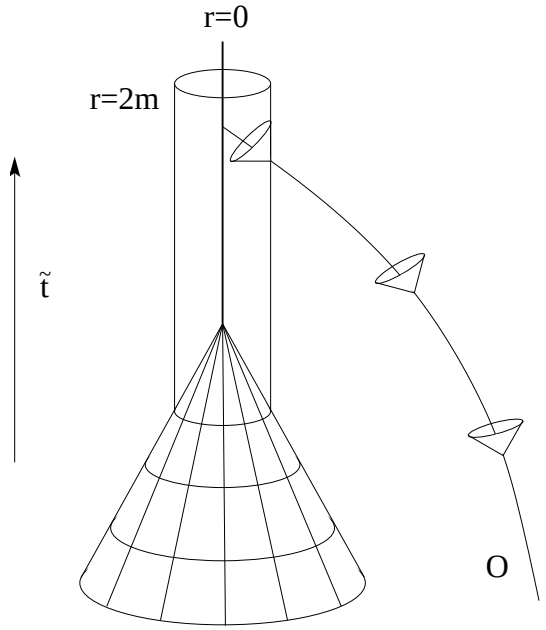


Figure 2.16: A spacetime diagram showing how a star collapses to form a black hole. The angle ϕ is suppressed. The stars radius decreases until it reaches the singularity at $r = 0$ and the event horizon is formed at $r = 2m$. The event horizon encloses the events which cannot be seen from the outside world. The light cones progressively tilt over as the approach the singularity and are parallel to event horizon at $r = 2m$, thus to show that this is the causal limit between the inside of the black hole and the outside.

The event horizon is a null surface at $r = 2m$ this represents the transition from the outside of a black hole, $r > 2m$ into the inside of the black hole, $r < 2m$. This is a point at which only one way traffic is possible into the black hole, towards the singularity at $r = 0$. The event horizon is a trapping surface which only allows matter and radiation to fall inwards and does not allow any matter or radiation to escape from it[2]. At the Schwarzschild radius $r = 2m$ an outgoing light ray cannot escape and is held by the gravitational field and dragged back towards the singularity at $r = 0$. This gives rise to the name black hole since not even light can escape leaving a black hole to any observer outside the event horizon. We can thus understand that the event horizon acts as a causal limits which separates the inside of the black hole from the outside.

A black hole's event horizon grows in size as the black hole grows. This is apparent by virtue of the fact that the event horizon occurs at $r = 2m$ and therefore as the mass increases, so does the radius the event horizon.

The singularity at $r = 0$ requires some further discussion and we need to address what it is. This is a real physical singularity and the spacetime curvature diverges at this point. General relativity fails to give a complete description of what happens at this point since spacetime itself breaks down. It is however hidden from the external world by the cloak of the event horizon of the black hole.

When considering quantum effects the "rules of censorship" applicable to the event horizon

[2]This is only for the classical situation and quantum effects can be taken into account which then alters the situation significantly.

become less appropriate. The reason for this is that it now becomes possible for a black hole to emit black body radiation due to the uncertainty principle, [Haw71]. It is this fact that forms the central basis of the next chapter and is discussed in more detail there.

There is another set of coordinates used to overcome the singularity at $r = 2m$ of the Schwarzschild metric of Eq. (2.64). This is called the Kruskal and Szekeres coordinates which is related to the Schwarzschild coordinate's in the following way,

$$\left.\begin{aligned} U &= \left(\tfrac{r}{2m} - 1\right)^{1/2} e^{r/4m} \cosh\left(\tfrac{t}{4m}\right) \\ V &= \left(\tfrac{r}{2m} - 1\right)^{1/2} e^{r/4m} \sinh\left(\tfrac{t}{4m}\right) \end{aligned}\right\} \text{when } r < 2m, \tag{2.66}$$

$$\left.\begin{aligned} U &= \left(1 - \tfrac{r}{2m}\right)^{1/2} e^{r/4m} \sinh\left(\tfrac{t}{4m}\right) \\ V &= \left(1 - \tfrac{r}{2m}\right)^{1/2} e^{r/4m} \cosh\left(\tfrac{t}{4m}\right) \end{aligned}\right\} \text{when } r < 2m, \tag{2.67}$$

where the coordinate patches have been split into an inside and outside section around the Schwarzschild radius. A good motivation for the introduction of these coordinates is given in [MTW73]. Applying the above transformation to the Eq. (2.64) one arrives at the following,

$$ds^2 = \left(\frac{32m^3}{r}\right) e^{-r/2m}(-dV^2 + dU^2) + r^2(d\theta^2 + sin^2\theta d\phi) \tag{2.68}$$

The function defining r in terms of v and u is given by,

$$\left(\frac{r}{2m} - 1\right) e^{r/2m} = U^2 - V^2.$$

The Kruskal-Szekeres metric is not singular at $r = 2m$ showing again that the singularity there in the Schwarzschild coordinates is just a coordinate singularity.

Insight can be gained into the Schwarzschild geometry by plotting lines of constant coordinates r and t on a UV grid. This is called a Kruskal diagram and is illustrated in Figure 2.17 (i). Lines of constant r produce hyperbolas, constant curves of $U^2 - V^2$, in the UV plane. The Schwarzschild radius $r = 2m$ is a straight line $V = \pm U$. The hyperbola

$$V = \sqrt{U^2 + 1}$$

corresponds to the value of $r = 0$. In a similar way the lines of constant t can be plotted by deducing

$$\begin{aligned} \tanh\left(\frac{t}{4m}\right) &= \frac{V}{U}, \quad r > 2m, \\ \tanh\left(\frac{t}{4m}\right) &= \frac{U}{V}, \quad r < 2m, \end{aligned}$$

from Eq. (2.66) and Eq. (2.67). Lines of constant t are straight lines through the origin of the U and V. The value $t = 0$ corresponds to the line $V = 0$ for $r > 2m$, whereas for $r < 2m$ it is the line $U = 0$. Region I of the Kruskal diagram with $U > 0, -U < V < U$ is covered by the Schwarzschild coordinates $-\infty < t < +\infty, 2m < r < \infty$. The entire Eddington-Finkelstein

coordinates $-\infty < v < +\infty, 0 < r < \infty$ is mapped into the part of the diagram with $V > -U$. That is the region through which the world line of the collapsing star moves and only that part outside the star's surface is relevant for spherical collapse.

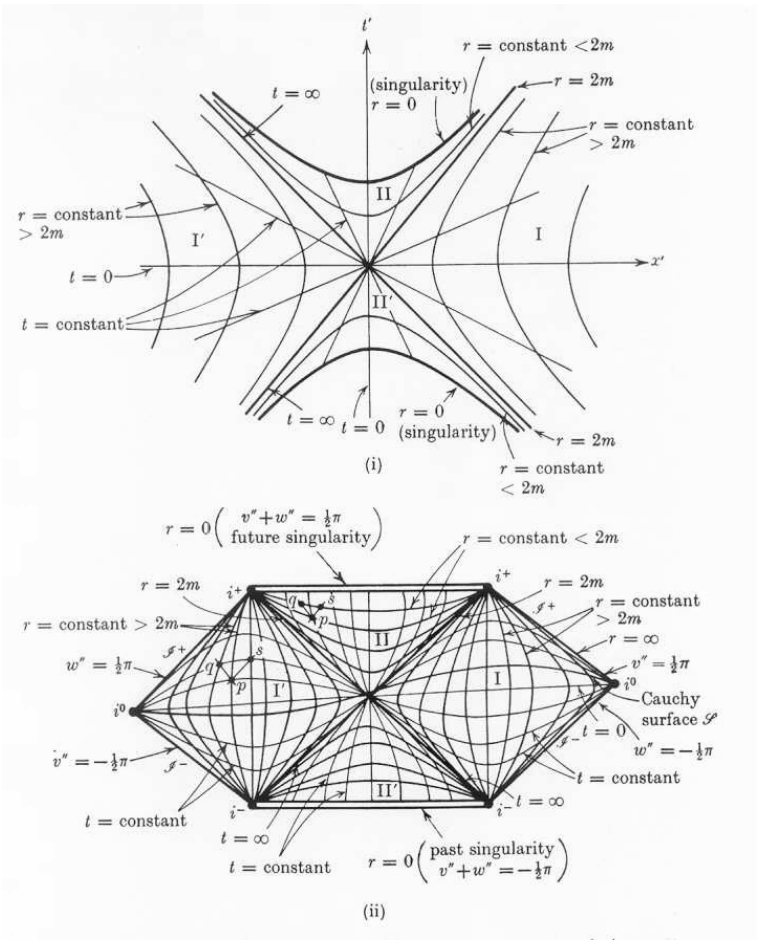


Figure 2.17: The maximal analytic Schwarzschild extension. The θ, ϕ coordinates are suppressed; null lines are at $\pm 45^o$. Surfaces $\{r = \text{constant}\}$ are homogeneous. (i) The Kruskal diagram showing asymptotically flat regions I and I' and regions II, II' for which $r < 2m$. (ii) A Penrose diagram showing the conformal infinity and the two singularities. This diagram is taken from [HE73].

It is possible to map a Kruskal diagram into a Penrose diagram by manipulating the coordinates in such a way that light rays propagate along 45^o lines and relabelling infinity to become a finite value. The Penrose diagram for the Schwarzschild geometry proves useful in understanding the global properties of spacetime. We define

$$U = \frac{\tilde{v} - \tilde{u}}{2}, \qquad V = \frac{\tilde{v} + \tilde{u}}{2}.$$

This relabelling just allows for the axis of UV to be rotated by 45^o, for the purpose of light rays moving along curves of constant u or v. Lastly introduce two new coordinates (u', v') and

(U', V') defined by

$$u' \equiv tan^{-1}(\tilde{u}) \equiv V' - U',$$
$$v' \equiv tan^{-1}(\tilde{v}) \equiv V' + U'.$$

This is illustrated in Figure 2.17 (ii) where we can now see the light rays move on the 45^o lines of the $U'V'$ plane. We can now map the infinite ranges of u and v to the finite range of $(-\pi/2, \pi/2)$ for u' and v'. The horizon $V = U$ maps into the same 45^o line in the $U'V'$ plane. The regions I, II, III and IV of the Penrose diagram, Figure 2.17 (ii), are isometric to the same respective labelled regions in the Kruskal diagram, Figure 2.17 (i).

This Penrose diagram of the Schwarzschild geometry contains a past and future null infinity at $\mathcal{I}^+$ and $\mathcal{I}^-$, respectively. While a timelike infinity are at i^+ for the future and i^- for the past. exist as well along with a spacelike infinity at i^0. The regions I and II are of the most significant from a physical perspective with regions III and IV being more a theoretical complement (analytic continuation) to the first two regions. Region I is asymptotically flat, and is thus very similar to Minkowski space in this region at a large radius. Beyond the event horizon at $r = 2m$ we have region II which is a closed trapped surface and thus anything in this region inevitably ends up falling to the future singularity.

Chapter 3
Entropy Bounds

3.1 Introduction

General relativity and quantum mechanics are the two foundational pillars in modern physics. Both these disciplines have proved their respective value in the realm of physics as being very secure although there formulation is completely independent of each other. A single theory that integrates these two paradigms still hasn't been developed completely. Investigating and solving the problems that occur when combining these two theories may prove to be a significant aspect in developing a completely unified theory.

One candidate for combining the theories is in the realm of gravitational instabilities. Quantum mechanics has no means for accounting for gravitational effects and thus doesn't have a clear cutoff at which these instabilities occur. Gravitational instabilities in general relativity, however, are well defined and result in the formation of black holes at sufficiently high energies in a given region of space. Black holes have a thermodynamic nature and their temperature can be determined by virtue of a semi-classical calculation. Black holes thus pose a natural and interesting setting, since it may lead to some new insights in which to explore the crossover between these two disciplines. The expected result of examining these two theories in this setting is that a cutoff will emerge for gravitational instabilities within quantum mechanics and this will have bearing on quantum gravity since we are dealing with very high energy scales and short length scales (Planck).

Section 3.1.1 reviews 't Hooft's way of combining aspects of general relativity and quantum mechanics within the setting of black holes, albeit in a non-rigorous way. This example is the key to introducing and conceptualising some of the aspects that will be explored in detail in this chapter. The expected result of arriving at a bound is achieved for the gravitational instability but the surprising character of this bound is that it is related to the area of the event horizon. This means that we are able to relate the number of degrees of freedom, or information, associated with the volume of a black hole to simply an area. This is reminiscent of a Hologram in which all the information of three dimensional object can be captured on a two dimensional surface. This conjecture is thus aptly named the Holographic Principle. The significance of this may be far reaching and requires further investigation.

The question we need to address in a broader context is how much information is required, at a fundamental level, to encode the universe. This lies at the heart of what the Holographic Principle suggests. We need to thus determine if this is a plausible conjecture within the well defined and rigorous foundation of general relativity while using sound quantum mechanical principles to steer us in our exploration in order to learn if the Holographic Principle could be a fundamental theory underlying physics.

This is what we attempt to address in this chapter. To answer this question the spherical entropy bound is introduced and is constructed by way of black hole thermodynamics. The spherical entropy bound is eventually refined to become the covariant entropy bound and is ultimately tested in a variety of settings. This chapter follows a similar outline presented in

[Bou02].

3.1.1 't Hooft Example

An example is reviewed by 't Hooft, [tH85], that allows you to derive the density of quantum states for a black hole. This takes into account the Hawking radiation, Section 3.2, of a black hole and its associated temperature, $kT_H = 1/8\pi M_{BH}$. There are two ways in which to do this. The simplest of which is to use thermodynamics but this may bring into question whether a black hole is truly in thermodynamic equilibrium since if energy is added to a black hole its mass and size increases appropriately, which in turn results in the temperature dropping. The second way is to compute the spectral density of black hole directly from the Hawking temperature. This requires time reversal invariance.

We have at our disposal the emission rate (the Hawking radiation intensity) and the capture probability, which can be characterised as the black hole's effective cross section. The cross section is approximated as

$$\sigma \approx 2\pi R^2 = 8\pi M^2, \tag{3.1}$$

where the value would be slightly larger for objects moving slowly near the black hole. The emission probability $W\mathrm{d}t$ for a given particle, in a given quantum state, in a large volume $V = L^3$ is,

$$W\mathrm{d}t = \frac{\sigma(\mathbf{k})v}{V} e^{-E/kT}\mathrm{d}t, \tag{3.2}$$

where $\mathbf{k}$ is the wave number characterising the quantum state, v is the particle velocity and E is its energy.

Now we assume that the process is governed by a Schrödinger equation. This means that there exist quantum mechanical transition amplitudes,

$$\mathcal{T}_{\text{in}} = {}_{\text{BH}}\langle M+E|\mathcal{T}|M\rangle_{\text{BH}} \otimes |E\rangle_{\text{in}} = {}_{\text{BH}}\langle M+E|\mathcal{T}|\text{Absorbed}\rangle, \tag{3.3}$$

$$\mathcal{T}_{\text{out}} = {}_{\text{BH}}\langle M| \otimes {}_{\text{out}}\langle E|\mathcal{T}^\dagger|M+E\rangle_{\text{BH}} = \langle \text{Radiated}|\mathcal{T}^\dagger|M+E\rangle_{\text{BH}}, \tag{3.4}$$

where the state $|M\rangle_{\text{BH}}$ represents black hole states with mass M, and the other states are energy eigenstates of particles in the volume V. This then illustrates a transition between an initial state. This is accomplished by virtue of the time evolution operator, $\mathcal{T}$, which is also used in the reverse process,$\mathcal{T}^\dagger$, where energy is radiated from the initial black hole state. Now rewriting these amplitudes using the Fermi Golden Rule, the cross section and emission probabilities can be written as

$$\sigma = |\mathcal{T}_{\text{in}}|^2 \varrho(M+E)/v, \tag{3.5}$$

$$W = |\mathcal{T}_{\text{out}}|^2 |\varrho(M)/V, \tag{3.6}$$

where $\varrho(M)$ stands for the level density of a black hole with mass M. $|\mathcal{T}_{\text{in}}|^2$ thus represents the probability of the level density $\varrho(M+E)$ being attained, while similarly $|\mathcal{T}_{\text{out}}|^2$ is the probability associated with the final state $\varrho(M)$ being achieved. The factor v^{-1} in σ is a kinematic factor, and the factor V^{-1} in W arises from the normalisation of the wave function.

In this instance we require PCT invariance but the parity transformation P and charge conjugation C have no bearing on our calculation of σ. We do demand time reversal, $\mathcal{T}^{-\infty} = \mathcal{T}^{\dagger}$, this means we have a unitary transformation between the $|\text{Absorbed}\rangle$ and $|\text{Radiated}\rangle$ states. A consequence of this allows us to relate $\mathcal{T}_{\text{in}}$ to be equivalent with $\mathcal{T}_{\text{out}}$. Dividing σ by W and using Eq. (3.2) one finds

$$\frac{\varrho(M+E)}{\varrho(M)} = e^{E/kT} = e^{8\pi ME}. \tag{3.7}$$

Bearing in mind that E is very small relative to M we can complete a first order Taylor expansion around M, $1 + \frac{1}{\varrho(M)}\frac{d\varrho}{d\varrho(M)}\delta E = 1 + 8\pi\delta E$, and then finally integrate this expression to give

$$\varrho(M) = e^{4\pi M^2 + C} = e^{S}, \tag{3.8}$$

where $S = \ln \varrho(M)$. If we let $A_0 = 4\ln 2$, remembering that all areas are in units of Planck length squared, then we can represent the above expression as

$$\varrho(M) = 2^{A/A_0}. \tag{3.9}$$

There are thus two accessible orientations, just as in a simple two dimensional spin system, and

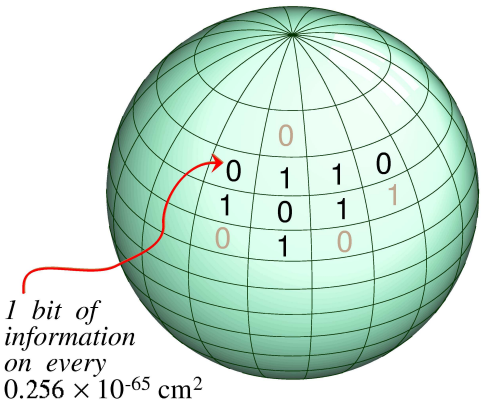


Figure 3.1: Information bits per unit area on a black hole horizon. Based on [tH85].

each of these two states occupies a unit of area, A_0. This then illustrates how you can imagine the horizon of a black hole having a spin-like degree of freedom for every surface element A_0, as shown in Figure 3.1. This seems to suggest a fundamental resolution for information and most importantly this resolution is per unit *area*.

3.2 Black Hole Thermodynamics

There is an unsettling argument called the No-Hair Theorem, discussed by Hawking amongst others, [Haw71] and [Haw72], that states that a stationary black hole is characterised by only three parameters. These are mass, charge and angular momentum. Consider a complex matter system such as a star, which would have many microstates associated with it, that collapses into a black hole. By virtue of conservation of charge, angular momentum and energy the

resulting black hole would be characterised uniquely by these parameters and would have only one associated microstate. In other words in this example we have thus gone from a very high entropy of a collapsing star to a black hole which has none.

This seems to be in blatant violation of the second law of thermodynamics which states that entropy should always stay the same or increase for an isolated system. Although it can be argued that there do exist instances where the entropy of a system may decrease over time, this is always only a fluctuation on a small scale and not a permanent and global effect, as is the implication from the above example.

Bekenstein was hesitant to accept such a violation and wanted to re-engineer the second law of thermodynamics to hold for the instance of a black hole as well. He used a theorem by Hawking called The Area Theorem of black hole's, [Haw71], which states that the area of a black hole's event horizon never decreases with time, i.e. $dA \geq 0$. The area is related to the mass of the black hole in the following way, $A = 16\pi M^2$ in $D = 4$. The constant increase in area of a black hole dictated by the Area Theorem is similar in character to that of the constant increase in entropy ensured by the second law of thermodynamics and an analog between the area and the entropy is thus possible. This allowed Bekenstein to conjecture, [Bek72], [Bek73] and [Bek74], that the entropy, or additional microstates, of the black hole is related with the area of the event horizon

$$S_{\mathrm{BH}} = \eta A = \frac{A}{4}, \tag{3.10}$$

where $\eta = 1/4$ for $D = 4$. Now that we have established a means to determine the entropy of a black hole we need to fix the second law. This was done by restating it as follows: The sum of the entropy of a matter system and entropy of a black hole system, always increases. This is known as the generalised second law, [Bek72], [Bek73] and [Bek74], and is given as

$$S_{\mathrm{total}} = S_{\mathrm{matter}} + S_{\mathrm{BH}}. \tag{3.11}$$

The implications of the generalised second law can best be illustrated through an example: Consider two system, one that has only matter entropy (there is only matter present in this system) and the other has only black hole entropy (there are only black holes in this system). These two system are well separated and non-interacting. The initial total entropy is the sum of the two systems,

$$S^{\mathrm{initial}}_{\mathrm{total}} = S_{\mathrm{matter}} + S_{\mathrm{BH}}. \tag{3.12}$$

If we now allow the two systems to interact until a new equilibrium is reached, we may find matter systems interacting with one another and turning into black holes and black holes similarly combining with other black holes. We now have a new system different to our initial systems.

If we now add together the entropy of the matter and black holes, this would produce a final total for the entropy. What the generalised second law prescribes is that this final entropy is

greater than or equal to the initial entropy,

$$S_{\text{total}}^{\text{final}} \geq S_{\text{total}}^{\text{initial}}. \tag{3.13}$$

The final element of developing a thermodynamic descriptions of black holes was to find an associated temperature. This was established through the findings of Hawking, [Haw74] and [Haw75], which proved an elementary consequence of quantum field theory when applied to fields living in a black hole metric. The manipulation of creation and annihilation operators, [Haw74] and [Haw75], showed that the states near the horizon are not truly vacuum, but actually contain a precisely computable density of particles, which are emitted as black hole radiation at a temperature

$$kT_H = \frac{1}{8\pi M_{BH}}, \tag{3.14}$$

where k is the Boltzmann constant.

3.3 Spherical Entropy Bound

We now are going to consider the consequences of the generalised second law for the process of a collapsing star. We require the following conditions,

- The matter entropy resides on a spacetime metric that allows for the formation of black holes. To ensure this we consider a metric that is asymptotically flat.
- The area A is defined to be a circumscribing sphere of the system. We require spherical symmetry and thus weakly gravitating systems.
- The time dependence of A should also be negligible and it should therefore not expand or collapse rapidly.

Let us consider an isolated star, which is composed of matter and we thus only consider the associated entropy of this matter or matter entropy (no black hole entropy to be accounted for). The total mass of this star is E, just less than that of a black hole. We now have a shell that closes in on this system, this shell is also composed of only matter entropy and has a mass that is just sufficient so that the star becomes a black hole. The process explained above is essentially a reaction we are setting into motion by adding two components of matter that have a combined critical mass to form a black hole for a given radius. This is known as the Susskind Process, [Sus95].

$$S_{\text{total}}^{\text{initial}} = S_{\text{matter}} + S_{\text{shell}} \tag{3.15}$$

We started with a entropy of S_{matter} added to it S_{shell} and we ended with a final entropy equal to that of a black hole, $A/4$.

$$S_{\text{total}}^{\text{final}} = S_{\text{BH}} = \frac{A}{4} \tag{3.16}$$

The generalised second law dictates that the final entropy is greater than the initial entropy and thus we have a bound that tells us that the most matter entropy we can have in a given region is less than or equal to $A/4$ of that region. This is known as the spherical entropy bound.

3.4 The Big Question

We can now attempt to address the question we posed in the introduction, but is re-structured and stated in the following way, *"How many degrees of freedom are there at the most fundamental level?"*. This is still the same question as before but uses the fact that information and degrees of freedom are related and interchangeable to alter the question so that it is in terms of the dimensionality of the Hilbert space.

Let us first look at what sort of answer we would get by addressing this question from the perspective of traditional Field Theory and then we can look at how the spherical entropy bound would do it. In order to evaluate these two view points on an equal footing we have to deal with a system that adheres to the conditions layed out in the spherical entropy bound. For this purpose we define a "fundamental system" as a region were the metric is spherically symmetric and not strongly time dependent.

It is important to note that the one alteration we have made to our initial question is that we are now actually looking at a region of space and not the entire universe. This is also something of a paradigm shift from conventional quantum mechanics in that we are not looking at a specific system but rather a given region.

3.5 Complexity according to Local Field Theory

To start with we consider the fundamental system in a well understood setting of a local quantum field theory on a classical background satisfying Einstein's equations. In general a quantum field theory consists of an oscillator at every point in space. Even something like a single Harmonic oscillator has an infinite dimensional Hilbert Space. Moreover, there are infinitely many points in any volume of space, no matter how small. Thus, the answer to our question, at this point, appears to be $N = \infty$. However, we have so far disregarded the effects of gravity altogether.

A finite estimate of the number of degrees of freedom is obtained by including gravity at least in a crude, minimal way. One might expect that distances smaller than the Planck length, $l_{\mathrm{P}} = 1.6 \times 10^{-33}$cm, cannot be resolved in quantum gravity. So let us discretise space into a Planck grid and assume that there is one oscillator per Planck volume, Figure 3.2.

The oscillator spectrum is thus discrete and bounded from below by finite volume effects. It is bounded from above because it must be cut off at the Planck energy, $M_{\mathrm{P}} = 1.3 \times 10^{19}$GeV. This is the largest amount of energy that can be localised to a Planck cube without producing a black hole. Thus, the total number of oscillators is V [3] and each has a finite number of states, n [4]. Hence, the total number of independent quantum states in the specified region is

$$\mathcal{N} = an^V. \tag{3.17}$$

[3] In Planck units.

[4] A minimal model one might think of is a Planckian lattice of spins, with $n = 2$

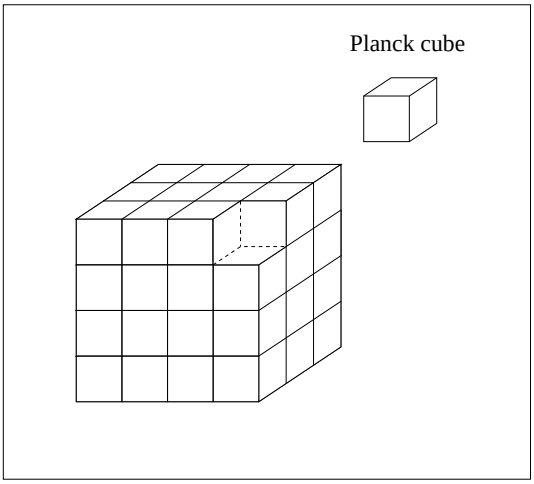


Figure 3.2: This is an illustration of a cubic volume of space which has been discretised into Planck cubes, each containing a single oscillator. The length of each side of the Planck cube is 1.66×10^{-33}cm.

The number of degrees of freedom[5] is given by

$$N = V \ln n \gtrsim V. \tag{3.18}$$

This result successfully captures our prejudice that the degrees of freedom in the world are local in space and related to the entire volume of the region considered.

3.6 Complexity according to the spherical entropy bound

Thermodynamic entropy has a statistical interpretation. Let S be the thermodynamic entropy of an isolated system at some specified value of macroscopic parameters such as energy and volume. Then e^S is the number of independent quantum states compatible with these macroscopic parameters. Thus entropy is a measure of our ignorance about the detailed microscopic state of a system. One could relax one of the macroscopic parameters by requiring that the energy only lie in a finite interval as opposed to being a specific value. This would then allow for more states and the entropy will be correspondingly larger.

To summarise the connection between thermodynamics and quantum statistical mechanics, we can rephrase the above and state the relationship in a more rigorously form: Thermodynamics results from coarse graining a more microscopic description so that states with similar macroscopic behaviour are lumped into a single thermodynamic state. The existence of thermodynamics will be taken to mean that a microscopic set of degrees of freedom exists whose coarse graining leads to the thermal description. More specifically we assume that thermodynamic entropy S implies that approximately $\exp(S)$ quantum states have been lumped into one thermal state.

The question at the beginning of this section was "How many independent states are required to describe all the physics in a region bounded by an area A?" Recall that all thermodynamic

[5]Where the number of degrees of freedom for a quantum mechanical system, N, is defined to be the logarithm of the dimension $\mathcal{N}$ of its Hilbert space $\mathcal{H}$ expressed as $N = \ln \mathcal{N} = \ln \dim(\mathcal{H})$

systems should ultimately be described by the same underlying theory, and that we are interested in the properties of this "fundamental system". We are now able to rephrase the question as follows: "What is the entropy, S, of the 'fundamental system', given that only the boundary area is specified?". Once we know how to answer this question, the number of states will simply be determined by $\mathcal{N} = e^S$, as explained earlier.

For the spherical entropy bound the entropy can be determined without any knowledge of the nature of the "fundamental system". The bound,

$$S \leq \frac{A}{4}, \tag{3.19}$$

makes reference only to the boundary area; it does not care about the microscopic properties of the thermodynamic system. A black hole that just fits inside the area A has entropy

$$S_{\text{BH}} = \frac{A}{4}, \tag{3.20}$$

so the bound can clearly be saturated with the given boundary conditions. Therefore, the number of degrees of freedom in a region bounded by a sphere of area A is given by

$$N = \frac{A}{4}; \tag{3.21}$$

the number of states is

$$\mathcal{N} = e^{A/4}. \tag{3.22}$$

We assume that all physical systems are larger than the Planck scale. Hence, their volume will exceed their surface area, in Planck units. The result obtained from the spherical entropy bound is thus at odds with the much larger number of degrees of freedom estimated from local field theory.

3.7 Which answer is correct?

The next section attempts to show why the field theory answer over counted the available degrees of freedom as a result of not properly taking into account gravitational effects.

In the field theory estimate the UV[6] cutoff tells us that we can have a maximal mass per Planck cube. As soon as this mass is exceeded, a black hole is formed. Thus if we would want to consider the formation of a black hole in a region of space, lager than a single Planck unit, we would have to count all the Planck cubes, R^3, in order to know what the associated critical mass, M, is for the formation of the black hole. This implies that there is a scaling relationship of $R^3 \sim M$ in the field theory which determines the point at which it becomes gravitationally unstable. In general relativity, however, the Schwarzschild solution tells us that the relationship for a black hole's formation should be $R \sim M$.

The discrepancy between these two answers is shown in Figure 3.3. Here it can be seen that the field theory solution allows for gravitationally unstable states as its scaling relationship diverges away from the linear relation demanded by the Schwarzschild solution. This disparity

[6]Ultra Violet or high frequency.

becomes more pronounced as we consider larger regions of space. The implication of this is that the field theory estimate includes states that are not accessible since they are at a gravitationally unstable level. Put another way this mean that while trying to excite the quantum field to these high energy states a black hole would result long before these states could be reached. From this we can conclude that not all the degrees of freedom that field theory apparently supplies can be used for generating entropy or storing information.

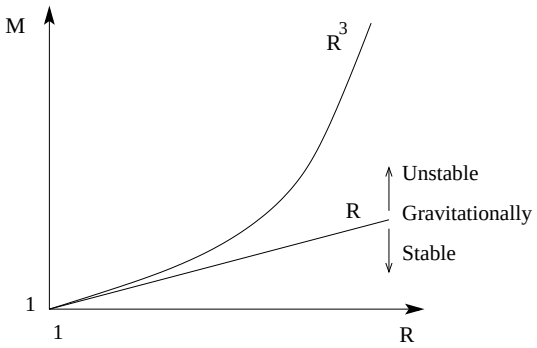


Figure 3.3: A graph depicting the relationship between mass, M, and radius, R. The straight line, M vs R, represents gravitational stability on or below this line as prescribed by the Schwarzschild solution. The Field theory has a M vs R^3 relationship which is above the straight line and thus in the region of gravitational instability.

3.8 Unitarity and the Holographic Principle

Imagine a pragmatic field theorist that still maintained that the previous argument was not sufficient to convince him. He would contend that we can still have a theory where we would be able to access all the additional high energy states at the cost of gravitational stability. We could now excite more than A/4 degrees of freedom, though we would have to jump into a black hole to verify that we succeeded. With this argument the field theorist manages to retain all the degrees of freedom given by the field theory.

If we are willing to accept that we are dealing with asymptotically flat space then we know that the physics in this region can be described by a scattering matrix (S-matrix). This then leads us to consider unitarity, which proves to be a compelling consideration.

Quantum mechanical evolution preserves information and it thus takes a pure state and evolves it to another pure state. If we consider a Hilbert space where we have a dimension of e^V and suppose that this region of space was then converted to a black hole. According to black hole entropy the region is now described by a Hilbert space of dimension $e^{A/4}$. The number of states would have decreased and it would be impossible to recover the initial state from the final state. This would violate unitarity, hence the space must have a dimension of $e^{A/4}$ to start with.

The consideration of unitarity in the presence of a black hole led 't Hooft, [tH93], and Susskind, [Sus95], to embrace a more radical "holographic" interpretation. They felt that this pointed to something more general than the argument followed for the specific case of the spherical entropy bound and there was an underlying principle of physics at play. This idea led to the initial formulation of the Holographic Principle: "A region with boundary of area A is fully described

by no more than $A/4$ degrees of freedom, or by no more than one bit of information per plank area."

3.9 Concluding Discussion on the spherical entropy bound

The spherical entropy bound gives us a first glimpse at what considerations a Holographic Principle may require and how we may begin to incorporate these developments. It does have several shortcomings in its current formulation, the most prominent of which is the strong conditions that have to be put into place before it becomes a working theory.

The settings in which the spherical entropy bound does have difficulty should also serve as a yardstick for newer refined versions of the spherical entropy bound. What is required is a more general and robust conceptualisation of the spherical entropy bound. This should then be tested in the specific areas where the spherical entropy bound failed and also be considered in more general and yet unexplored settings. A more generic and rugged formulation is the next step.

3.10 Spacelike Bound

In light of the spherical entropy bound we could theorise that a more general bound is at play. This more general theory would still support the spherical entropy bound in a more specific setting. The most logical extension of the spherical entropy bound is to formalise the bound as being the boundary of a given area. This then forms the entropy bound.

The restrictive nature of the spherical entropy bound is due to the fact that it has a set of conditions associated with it. Remembering the conditions are an asymptotic structure, gravitational stability and spherical symmetry. By not including these conditions we achieve a more general bound which we will call the spacelike entropy bound. This bound conjectures that the entropy in a given spatial region will not exceed the area of that regions boundary. This is illustrated in Figure 3.4.

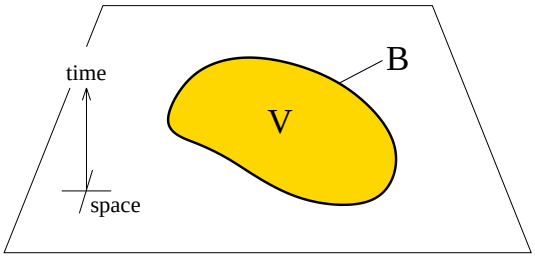


Figure 3.4: A hypersurface at an equal time with a spatial region V bounded by B. The spacelike entropy bound conjectures a means to relate the entropy in the spatial region, V, to the area of its boundary, B, but fails. Taken from [Bou02].

All the the entropy, S, of a spatial volume, V, is given by $S(V)$. While the area, A, of a region, B, bounding the volume, V is given by $A[B(V)]$. This produces the relation

$$S(V) \leq \frac{A[B(V)]}{4}. \tag{3.23}$$

This particular construct falls short on several fronts. This is highlighted in the following counter examples which will be reflected upon in later discussions.

3.10.1 Closed space

This is the simplest and most obvious example of how the spacelike entropy bound fails. Consider a closed spacelike hypersurface on a manifold $\mathcal{M}$. Let us call the entire volume of the space $\mathcal{V}$ this includes a non-zero matter system, $S_{\text{matter}}(V) = S_0 \geq 0$, that doesn't completely occupy the volume. Now let us consider a compact region Q which doesn't include the matter system. We now have a situation where the bound B of V coincides with the boundary of Q. This fact can be exploited, since there is no clear way of distinguishing the area inside the compact region Q and the outside bounded region $B(V)$. The area of Q can be contracted to a point. This leads to a situation were we have $S_{\text{matter}}(V) \geq A[B(V)]$ and the space like entropy bound, Eq. (3.23), is violated. There is an illustration of this scenario in Figure 3.5.

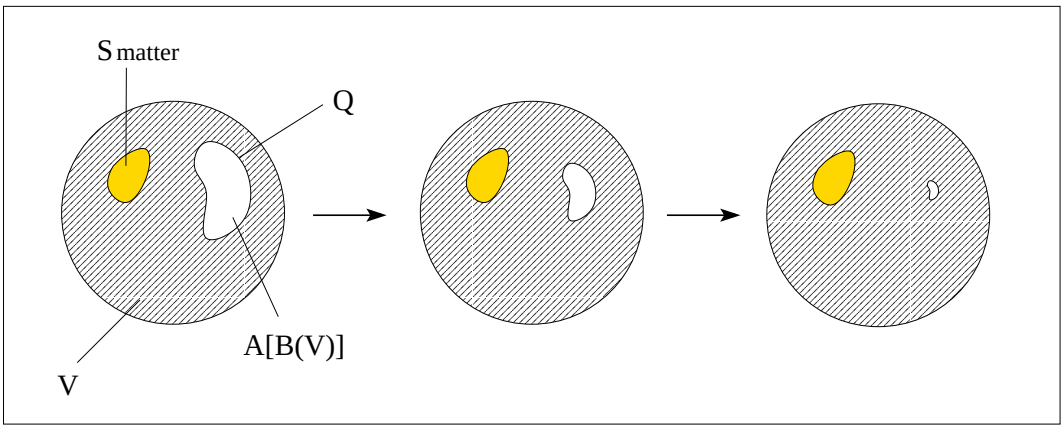


Figure 3.5: This is an illustration of the violation of the spacelike bound on a closed space. The unshaded area inside the compact region Q, $A[B(V)]$, is reduced to a point, from left to right in the figure. The $S_{\text{matter}}(V)$ is still contained in the bounded region of $B(V)$ but the area is being reduced to zero, resulting in a violation, $S_{\text{matter}}(V) \geq A[B(V)]$.

3.10.2 The Universe

Assume a flat, isotropic and homogeneous universe expanding with time. If we pick a homogeneous hypersurface of equal time, V, its entropy be can be captured as an average entropy density, σ, on this hypersurface. This results in

$$V = \frac{4\pi}{3}R^3, \quad A[B(V)] = 4\pi R^2, \tag{3.24}$$

since we are dealing with flat space. The entropy in the volume V is given by

$$S_{\text{matter}}(V) = \sigma V = \frac{\sigma}{3}RA = \frac{\sigma}{6\sqrt{\pi}}A^{3/2}. \tag{3.25}$$

Recalling the spacelike bound, Eq. (3.23), we can see that if we consider a large enough radius,

$$R \geq \frac{3}{4\sigma}, \tag{3.26}$$

it would violate the spacelike bound, [FS98].

3.10.3 Collapsing star

Consider a star with an initial entropy of S_0 that collapses to form a black hole after burning out. If we track its path as it falls through its own horizon, it is known from solutions of this type, [MTW73], that its radius eventually becomes zero at the singularity and thus its surface area does as well. The second law of thermodynamics requires that the enclosed volume considered would still be at least equal to the initial entropy of the star, S_0, but this would exceed $A/4$ as $A \to 0$ where we approach the singularity. This example is set in a strong gravitational field and the conditions of the spherical entropy bound would thus exclude it from being applicable.

3.11 Covariant Bound

As we have seen the spherical entropy bound is very useful in introducing the idea of entropy bounds but is very restrictive and requires several strong conditions for its construction. A more generic form of the spherical entropy bound is required that is applicable for a cross section of spacetime scenarios, this will be a constructive step in trying to substantiate the conjecture of the Holographic Principle. While attempting to generalise the spherical entropy bound a natural extension was formulated in the shape of the spacelike entropy bound but had serious inadequacies which culminated in the violation of the entropy bound, Section 3.10.1. A new perspective, that departs form the previous bounds, is required to develop a covariant construct. This requires a way in which to bound a region of spacetime so that considerations of the metric and the matter contents of a region of space are adhered to.

The work of Susskind, [Sus95], produced a way in which one could map the horizon of a black hole onto a holographic screen via the use of light rays and the focusing theorem, Section 3.12.1. This gave a result that preserved the nature of the entropy bounds we have introduced thus far but was expressed in a slightly different context, $\frac{\triangle S}{\triangle A_{\text{Projected}}} = \frac{1}{4}$. Fischler and Susskind, [FS98], realised there was something significant about the light rays introduced in the mapping of black hole to screens and developed an application in FRW cosmological spacetime that had a lightlike nature. The idea was that a spherical region considered at an early time after the big bang could be bounded by only considering the entropy density passing through a portion of the past directed light cone.

This, while solving many of the problems associated with the spacelike entropy bound which had also been applied in this setting, still suffered shortcomings. This did, however, mark a shift in development and the light cone was now introduced as the screen on which the entropy was projected and taken into account. The covariant entropy bound by Bousso, [Bou99], was a refinement and generalisation of the Fischler-Susskind proposal. Bousso formalised a way in which the lightsheets could be constructed and how the entropy could be calculated on a lightsheet bounding a region. It can be used in a diverse range of settings with regard to spacetimes and regions.

The *covariant entropy bound* states that *the entropy on any light-sheet of a surface B, will not exceed the area of B*:

$$S[L(B)] \leq \frac{A(B)}{4}. \tag{3.27}$$

The shift in perspective between the covariant entropy bound and the spacelike entropy bound

are important to note. They are comprised of two main differences the first of which is that in the spacelike entropy bound we have a volume V which is bounded by a region B, while in the covariant entropy bound we have a bounded region B that is enclosed by a lightsheet L. Secondly, and most significantly, V is a spacelike surface while L is a null hypersurface.

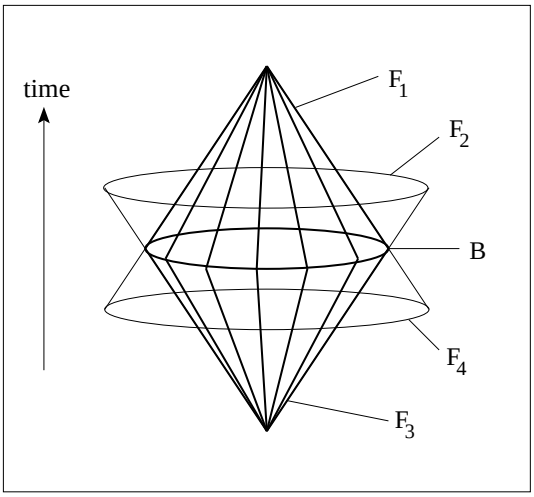


Figure 3.6: The four null hypersurfaces orthogonal to a spherical surface B. The two cones F_1, F_3 are the only lightsheets since they have negative expansion. The covariant entropy bound states that the entropy on each light-sheet will not exceed the area of B. The other two families of light rays, F_2 and F_4, have a positive expansion with no caustics resulting in an cross-sectional area which is increasing. They are not light-sheets. The entropy contained in F_2 and F_4 is not related to the area of B and thus serves little purpose. Based on illustration from [Bou02].

In order to understand the covariant entropy bound we will first examine the lightsheet formulation, Section 3.11.1, and then how we can calculate the entropy on the lightsheet, Section 3.11.3. Given plausible conditions for the energy and entropy relationship the validity of the formulation of the covariant entropy bound is strengthened but doesn't seem to have a fundamental footing. Similarly there is no comprehensive motivation for the use of lightsheets except for the historical formulation that pointed in this direction. The elements that contributed in the development in this direction are the success of the lightsheet formulation and the properties of the Raychaudhuri equation. We explore these issues in later sections.

3.11.1 Lightsheet formulation

The significant difference to other entropy bounds exhibited by the covariant entropy bound is that it relates a spacelike region with its associated adjacent lightlike region. This is done through the use of a lightsheet for a given region.

We need to establish how these lightsheets are related to spacelike surfaces or regions. This is fortunately done easily since any region of space B has exactly four orthogonal null (lightlike) directions associated with it. These four sets of light rays are guaranteed by virtue of the Lorentzian geometry and are independent of the shape and location of the selected region, B. These are light cones that are referred to as future directed, ingoing or outgoing, and past directed, ingoing or outgoing. The description of the above set of lightsheets is shown in Figure 3.6.

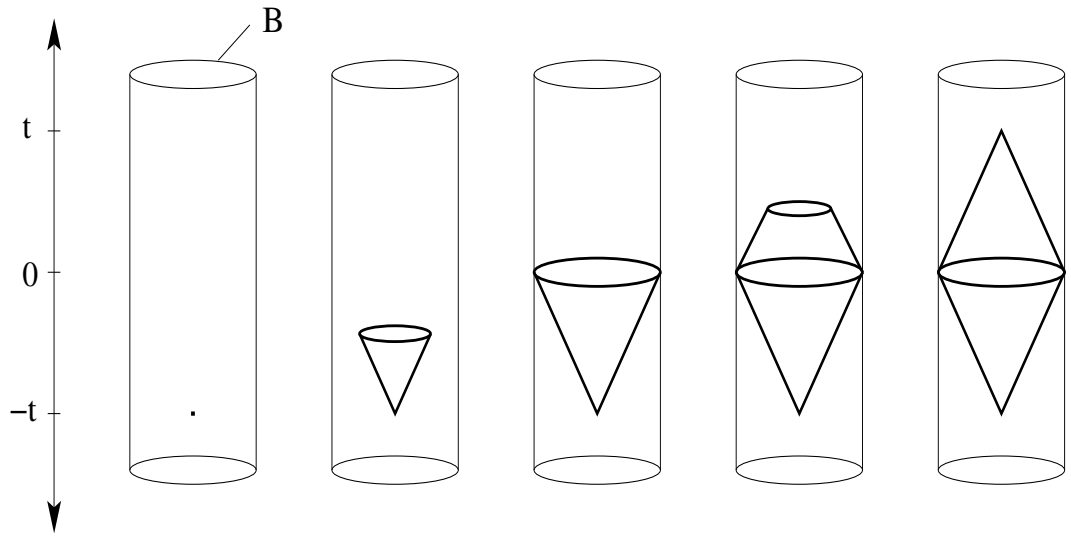


Figure 3.7: Inward going lightsheet formation as a function of regular time intervals. The cylinder depicted is the bounded region, B, and the formation of the lightsheet begins on the left and progresses to the right where the entire inward going lightsheet is shown for past, $-t \leq 0$, and future, $0 \leq t$, directions. One arrives at the lightsheets F_1 and F_3 illustrated in Figure 3.6.

In order to get a better conceptual feel for how the ingoing, past and future, lightsheets would be constructed, consider the illustration in Figure 3.7. Here a spherical room is lined with mirrors and has a single light bulb in the centre of the room. The light bulb is flashed at a time in the past, $-t$, as the light rays progressed outward from the centre of the room towards the mirrored walls, a light cone forms. When the rays reach the mirrored walls, $t = 0$, they are reflected inwards to form yet another light cone and terminate when they reach the centre of the room at t_1. These two light cones are the ingoing past, $-t$ to $t = 0$, and ingoing future, $t = 0$ to $+t$ directed lightsheets. One can now also line the outside of the spherically shaped room with mirrors and place an array of light bulbs on the limit of the spherically symmetric universe (the room is assumed to be in the centre of the universe) and then go through the same procedure, Figure 3.8, to form the past outward and future outward directed families of light rays.

We only require one of these lightsheets, for our purposes, but need a reliable and consistent way to choose it. Two of the candidates discussed above are unsuitable for bounding a region. They are expanding outwards and thus only succeed in capturing an arbitrarily large entropy associated with the infinite unbounded space outside the region being considered. We can explain this from perspective of the example given above where both the past and future outward directed family of light rays are reflected at the bound of the region and thus don't even enter the inside of the spherical room, which is the region of interest to us.

The labels of inward and outward directed lightsheets are awkward and clumsy. Lets not forget the problem that developed for the spacelike entropy bound in a closed space, Section 3.10.1, where it became difficult to distinguish between the correct bounding area simply because of a poorly defined concept of inside and outside. We must introduce a local condition that specifies how to do this. A general way in which we could do this is if we considered the area of the lightsheets, A, at a given time, $t = 0$, and again at a later time, t, with area, A'. If the area

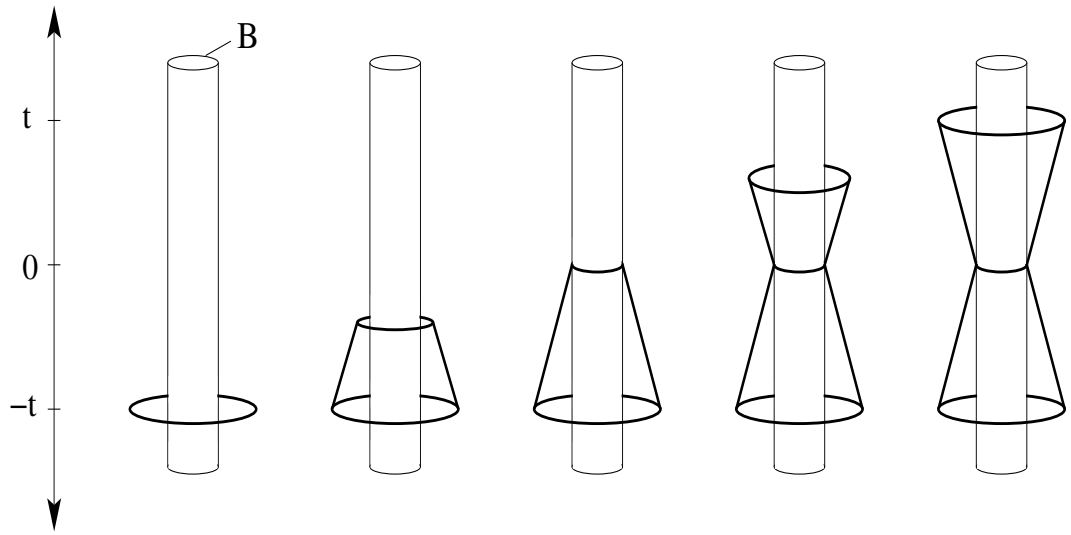


Figure 3.8: Outgoing family of light rays as a function of regular time intervals. The cylinder depicted is the bounded region, B, and the formation of the family of light rays begins on the left and progresses to the right where the entire outward going family of light rays is shown for past, $-t \leq 0$, and future, $0 \leq t$, directions. One arrives at the lightsheets F_2 and F_4 illustrated in Figure 3.6. It is clear that there is once again no direct relationship between the cross-sectional area of this family of light rays and the bounded area within B.

has decreased, $A' < A$, then we have an inward moving lightsheet and conversely if the area increased we can regard this lightsheet as outward moving.

The contraction condition is used as a tool to achieve the desired distinction of inward and outward moving lightsheets discussed above. This can be defined in a strict mathematical sense and is called the contraction condition. It demands a non-expanding nature of the lightsheets in order to be considered ingoing. It can be understood by considering an array of infinitesimally neighbouring light rays spanning a surface area $\mathcal{A}$.

The contraction condition for lightsheets is defined in a Lorentzian spacetime by firstly considering a convex closed surface B of codimension one and area A. We now construct four orthogonal geodesics intersecting B. These produce four null directed hypersurfaces. Considering that a length unit dl is meaningless in the context of light rays we need to parametrise the lightlike direction in terms of an affine parameter $d\lambda$. With this in place we are able to establish an associated direction with each of the lightsheets and consequently determine what is ingoing and outgoing.

If we follow each geodesic an infinitesimal portion along the affine parameter, $d\lambda$, to one of the two sides of B we produce an area, A'. If $A' < A$ the chosen direction is inside or ingoing while if $A' > A$ the direction is outside or outgoing. There is conceivably also the scenario where there is locally neither an expanding nor contracting direction. In this instance both members are considered inside. One also gets degenerate instances where there can be three or four ingoing directions. This condition can be applied to each infinitesimal surface element separately or in other words locally. It also applies to both closed and open surfaces.

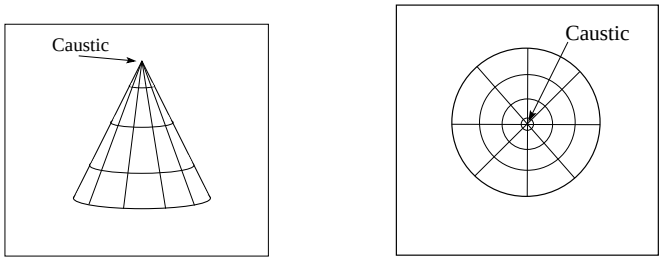
Casting the above into a mathematical form, we first define

$$\theta(\lambda) \equiv \frac{d\mathcal{A}/d\lambda}{\mathcal{A}}. \tag{3.28}$$

where λ is an affine parameter for the light rays generating the lightsheets F_i and we assume that λ increases in the direction away from B. We distinguish λ_0 as being the value of λ on boundary B. Then we are able to give the state that the contraction condition has to satisfy

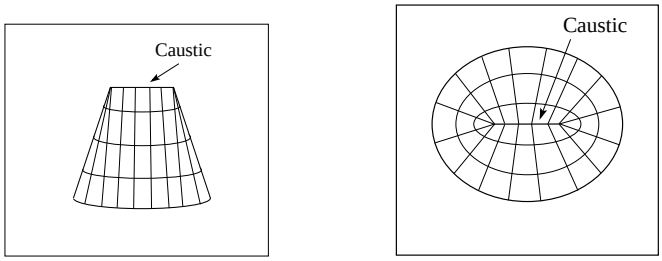
$$\theta(\lambda) \leq 0 \tag{3.29}$$

in terms of our definition. The expansion, θ, of a family of light rays is discussed in detail in Section 3.12.1.



(a) side view (b) top view

Figure 3.9: Caustic termination for simple spherical symmetric matter distribution. The caustic is simply the pinnacle of the light cone, this may not always be the case as illustrated in a more complex instance below.



(a) side view (b) top view

Figure 3.10: Caustic termination for a more complex symmetric matter distribution. This turns out to be a region at the top of the cone that pinches off in a straight line.

The last thing we now need is a way in which to know when to terminate the lightsheet. This is to prevent the same problem highlighted in the example of the outgoing lightsheets that end up covering an infinite area and have an arbitrarily large entropy. We want to be strict about imposing a bound to the region the lightsheet covers. By virtue of the contraction condition $\theta(\lambda_0) \leq 0$, which is guaranteed by Raychaudhuri's equation Section 3.12.1, the expansion can

only decrease. Positive expansion along the ingoing lightsheet only occurs when the neighbouring light rays intersect each other at caustics. This is where the lightsheet is then terminated. These regions of termination at the caustics may be a single point, for a spherical region given in Figure 3.9, but isn't limited to this instance. It may prove to have a far more complex region of termination along the caustic, Figure 3.10, by virtue of the shape of the bounding region and the matter density distribution.

A modified lightsheet that adheres less strictly to the contraction condition, was introduced in [TE99] and [FMW00] which both prove to simplify the calculation of the lightsheet. The weakened contraction condition allows for the termination of the lightsheets at any point where the light rays intersect each other, as long as they originated from the same bounded region B, and is thus not only limited to when *neighbouring* light rays intersect.

3.11.2 Bousso notational convention for lightsheets in Penrose diagram

There are three instances of lightsheet types that may be generated due to the dynamics of the surrounding spacetime geometry. Bousso, [Bou99], introduced his own notation that proves useful in identifying each lightsheet type within the context of Penrose diagrams.

Every point in a Penrose diagram represents a sphere. We may thus consider the boundary, B, as being a point around which is centred the two null inward directions of the lightsheets. These may either be future and past or a combination of these two. We depict this by drawing legs for the relevant direction shown in Figure 3.11.

A simple flat Minkowski spacetime would dictate a normal sphere with one past and one future directed lightsheet. More dynamic geometries may, however, result such as in an expanding universe where a sphere would have two past directed lightsheets; these surfaces are called anti-trapped. On the other hand a sphere in a collapsing universe or inside a black hole would result in two future directed lightsheets; these surfaces are in turn called trapped.

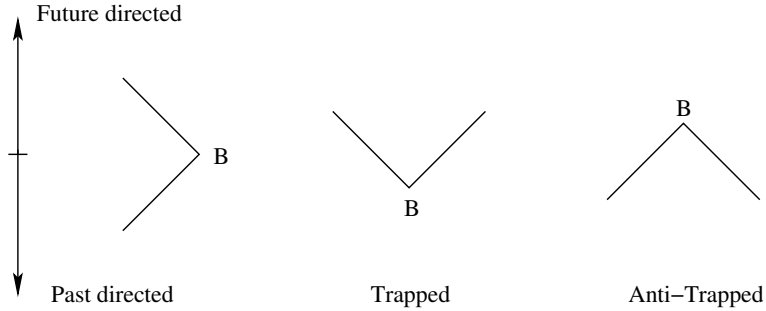


Figure 3.11: A figure showing the respective wedge notations of Bousso to easily identify the lightsheet of a bounded sphere, B, within a Penrose digram. The legs radiating outwards from B are either future or past directed. The first illustration on the left shows an instance where the two lightsheets are both future and past directed. The second schematic is called a trapped lightsheet and is where both lightsheets are future directed. These occur in strongly gravitating geometries (eg. Big crunch). The last depiction on the right shows an anti-trapped case where both lightsheets are past directed. These occur in a rapidly expanding geometry (eg. Big bang).

3.11.3 Entropy on Lightsheets

The question we are now confronted with is how do we define entropy on a light-sheet? The short answer is that there is no big difference in how you would define the entropy from an ordinary matter system and that on a lightsheet, there is only a slight shift in perspective. The entropy on a lightsheet is measured in light cone time and not a single instance in time. This means that you are literally taking a snap shot of the matter system. The entropy is still computed either thermodynamically or statistically by taking the logarithm of the number of quantum states. You can thus think of calculating the entropy on a lightsheet as riding a light ray and counting the numbers of states of the system as they travel through a region in space.

A problematic scenario arises when a lightsheet intersects only a portion of an isolated matter system. Flanagan, Marolf and Wald, [FMW00], made a statistical definition by which they demanded that the long wavelength modes which are not fully contained on a lightsheet not be included in the entropy. In a cosmological setting this is easily avoided since the entropy can be well approximated as a continuous fluid.

Up until now we have only dealt with entropy on a lightsheet which is a fixed null hypersurface. As with the example of an isolated thermodynamic system. In reality we must deal with different microstates which correspond to different distributions of energy. This has an effect on the geometry of the lightsheet. Although this is a small effect on average the lightsheet varies with the state of the system. On large scales, as in Cosmology, we can calculate the lightsheet as an average geometry and can estimate S_{matter} while holding the lightsheet geometry fixed. In a more general setting we can only hold B fixed and not the lightsheet of B. We must thus account for the fact that different configurations contributing to the entropy lead to macroscopically different future lightsheets.

There is no way in which the covariant entropy bound can account for the semi-classical Bekenstein entropy of a black hole since it is formulated only for matter systems in a classical geometry.

The covariant entropy bound is tied most naturally to a regime of approximately classical spacetimes where no quantum effects are taken into account. In order to clearly understand the workings of the covariant entropy bound we examine it in a setting that allows only for physically real scenarios. The null energy condition is thus enforced, this disallows negative energy but not a negative cosmological constant. Unlike other forms of negative energy a negative cosmological constant cannot be used to cancel the gravitational field of an ordinary thermodynamic system and therefore poses no problem for the holographic principle.

Quantum effects can violate the above energy condition. Casimir energy can be negative but its relation between duration, size and magnitude of such variations are severely constrained. Even where they do occur their gravitational effects can be over-compensated by those of ordinary matter. No counter example to the covariant entropy bound using quantum effects in an ordinary matter system has yet been constructed.

There does exist a contentious debate about whether quantum fluctuations do have an effect on the geometry of the lightsheet. The crux of this debate is centred on examples given by Lowe, [Low99], in which a black hole is fine tuned to violate the covariant entropy bound, Figure 3.12, but Bousso, [Bou00], contends that this scenario still holds, Figure 3.13, and Lowe does not

take full account of the quantum fluctuations. There are alternative formulations by Smolin in particular, [Smo01], that try to generalise the covariant entropy bound to a quantum gravitational setting.

The covariant entropy bound relies on geometric concepts such as area and orthogonal light rays and is thus developed to only really apply in classical spacetime but still has bearing on quantum gravity in terms of its applications in black holes and the AdS/CFT correspondence, [Mal98].

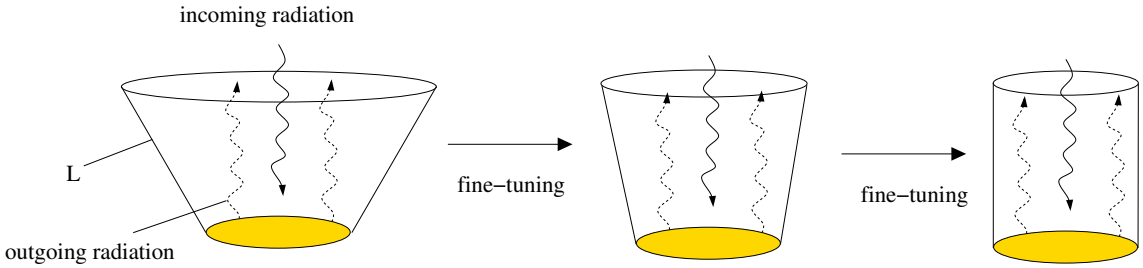


Figure 3.12: In Lowe's argument, [Low99], an evaporating black hole can be fine tuned to be in equilibrium with the incoming radiation. This would allow the future directed outgoing lightsheet of an area of the black hole to have vanishing expansion and thus continue to generate the horizon. This would allow for an unlimited amount of ingoing radiation entropy to pass through the lightsheet and ultimately violate the covariant entropy bound. The fine tuning is shown from left to right in the diagram of the outgoing lightsheet until it has vanishing expansion.

3.11.4 Summary

To summarise what we have learnt about the covariant entropy bound we begin with any D-dimensional Lorentzian space time M and introduce the definition of the covariant bound, [Bou02]: *Let $A(B)$ be the area of an arbitrary $D-2$ dimensional spatial surface B (which need not be closed). A $D-1$ dimensional hypersurface L is called a lightsheet of B if L is generated by*

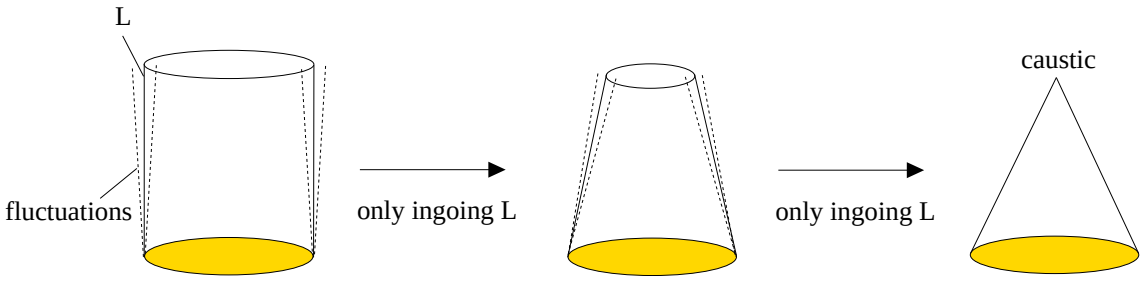


Figure 3.13: Bousso's counter, [Bou00], to Lowe's argument, [Low99], is that there will occur small fluctuations in the energy density of the radiation. This leads to the fluctuation of the lightsheet and if positive it must terminate while if it only fluctuates and never becomes positive then it will be negative on average. Within a finite affine parameter we will then see a termination of the lightsheet either by it becoming positive or averaging a negative expansion.

light rays which begin at B, extend orthogonally away from B, and have non-positive expansion,

$$\theta \leq 0, \tag{3.30}$$

everywhere on L. Let S be the entropy on any lightsheet of B. Then

$$S \leq \frac{A(B)}{4}. \tag{3.31}$$

It is useful to give a step by step account of how to implement the covariant entropy bound. This will lay the foundation for applying and testing the bound, as it will be demonstrated in Section 3.13.

- The first step is to determine the spatial region under consideration. This will be a $D-2$ dimensional surface B which will have an associated area $A(B)$. This area will also have four sets of light rays radiating orthogonally out from it, as depicted in Figure 3.6.
- The next step is to establish which of the light ray families we are going to use in our analysis of the covariant entropy bound. This requires that a spacetime metric is given for the region under consideration. There will be at least two light ray families of interest for which the expansion θ is negative. It is these sets of light ray we are trying to find.
- A lightsheet is an instance where one of the non-expanding families of light rays, F_j, reaches a caustic (a place where the light ray intersects with neighbouring light rays). This was illustrated in Figure 3.9. A lightsheet $L(B)$ is a null hypersurface of $D-1$ dimensions.
- Now determine the entropy as a function of the lightsheet, $S[L(B)]$. This was described in Section 3.11.3.
- The covariant entropy bound states that $S[L(B)] \leq \frac{A(B)}{4}$. Having established what the area $A(B)$ is for a spatial region in the first step and what the entropy $S[L(B)]$ is for each of the lightsheets related to that area in the fourth step, allows us to see if the bound holds.

The first three steps are concerned with selecting the lightsheets of the area of interest. In simple geometries one can determine the lightsheets simply by inspection. For a systematic analysis of more complex geometries we would require the geometric tools introduced in the next Section.

As long as we have a classical geometry with a physically realistic matter system, the lightsheet construction is well defined. The null energy condition, Section 3.11.3, is also important in formulating the covariant entropy bound successfully. This allows us to make the bound testable in real settings with the use of experimental data.

3.12 Dynamics of Light-Sheets

There must be an underlying thread that draws entropy and lightsheets together. This is what we aim to examine in this section. There is a simple line of reasoning that can be applied that will show the link between entropy and lightsheets. It begins by realising that entropy costs energy and energy, in turn, focuses light. Ultimately it is this focusing that leads to caustics being formed and prevents lightsheets from going on forever.

The Raychaudhuri equation and the focusing theorem is introduced in the next section to address the issue of how the lightsheets relate and reacts to an energy distribution.

In a significant paper by Flanagan, Marolf and Wald (FMW), [FMW00], it is shown that entropy can be approximated by a local flow of entropy density with plausible assumptions on the relation between energy and entropy.

Lastly we look to see if the covariant entropy bound can be reduced to some simpler more restrictive entropy bounds already discussed. This will establish the conditions under which it is possible to directly relate the covariant entropy bound to other bounds and the generalised second law of thermodynamics.

3.12.1 Raychaudhuri's equation and the focusing theorem

Expansion, shear and twist locally characterise a family of light rays that generate a lightsheet and in order to understand how these elements are affected by matter we need to evaluate the Raychaudhuri equation. This is explained in detail in [Bou02] but a overview is given below.

We begin by choosing one of the four light ray families, 3.6. Then the induced metric is established for the bounding surface, B. Information for the *expansion*, θ, *shear*, σ_{ab}, and *twist*, ω_{ab}, of the family of light rays, L, is determined by calculating the null extrensic curvature [Bou02]. At this point we can establish whether the expansion on the chosen lightsheet is non-positive and if not then it can be discarded and another lightsheet can be selected, remembering we are only looking for ingoing or non-positive expanding light rays.

The Raychaudhuri equation describes change in expansion along the light ray and is important in understanding the dynamics of the light ray

$$\frac{d\theta}{d\lambda} = -\frac{1}{D-2}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - 8\pi T_{ab}k^a k^b, \tag{3.32}$$

where $k^a = dx^a/d\lambda$. The twist vanishes in the above equation when considering a surface-orthogonal family of light rays, [Wal84]. The null energy condition is assumed through out, Section 3.11.3, and results in the term, $-T_{ab}k^a k^b$, being non-positive which in turn ensures that the right hand side of (3.32) is always non-positive. This gives rise to the following inequality

$$\frac{d\theta}{d\lambda} \leq -\frac{1}{D-2}\theta^2, \tag{3.33}$$

which forms the basis for the focusing theorem. This illustrates that if we consider a negative expansion, θ, we have a locally vanishing cross-sectional area until the light rays eventually terminate at the caustic point (intersection of neighbouring light rays).

There are two key points we can conclude from the discussion thus far. The first being that lightsheet encountering any form of matter or a positive stress energy tensor will result in the focusing of light rays according to Raychaudhuri's equation, (3.32). Secondly knowing that entropy relates to energy, which we have just established causes light rays to focus, we can conclude that entropy indirectly is associated with the focusing of light rays. It is important to note that the rate of the focusing is quick enough to ensure that caustics form before the covariant entropy bound can be violated but still slow enough to allow the bound to be saturated.

3.12.2 The covariant entropy bound formulated as conditions

The underlying mechanism of the covariant entropy bound is the interaction between entropy and energy. Entropy requires energy and it is this energy which focuses the lightsheets to produce the caustics of a bounded region in the setting of the covariant entropy bound. Thus the critical link between entropy and energy is what determines how successful the covariant entropy bound ultimately is. Quantitatively this relationship depends on the details of the specific matter systems and cannot be calculated in general.

Many situations exist where entropy can be approximated by the local flow of entropy density. A key step in resolving this problem is introduced by Flanagan, Marolf and Wald, [FMW00], who proposed that the covariant entropy bound is always satisfied if certain conditions apply to the entropy and energy density. This was successfully proved under two separate sets of conditions.

The first set of conditions are as difficult to prove as the covariant entropy bound itself. It does, however, shed light on the connection with the Bekenstein bound. The process under this set of condition is to construct the lightsheet, find the endpoints and then calculate the entropy by making an analysis of the modes. This set of conditions can be seen as a reformulation of the covariant entropy bound in terms of conditions.

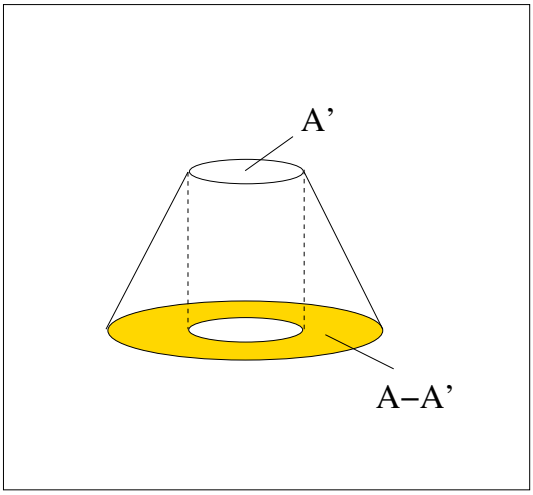


Figure 3.14: This figure shows a partial light cone that does not proceed to a caustic and is a graphical representation of the strengthened form of the covariant entropy bound. The shaded region at the base of the cone is the bounded area of the light cone, $A - A'$. The light cone doesn't bound the area A' and therefore cannot relate to the entropy in that region.

The second set of conditions dictates that there is a locally defined energy and entropy densities. The entropy content of a spacetime can be represented by a fluid approximation, which can be verified. A further verification is required to check the bound is intact for each of the lightsheets.

It must be remembered that neither of these sets of conditions have a fundamental law of physics underlying them. They do manage to strengthen the covariant entropy bound since they manage to exclude a large class of counter examples. The most useful finding, from a practical perspective, is that the second condition delivers a shortcut bound for cosmological examples. Bousso, [Bou02], raises concerns that the broad nature of these conditions may overlook some

underlying subtleties of the covariant entropy bound.

We now proceed to state these conditions precisely as two theorems in a summarised form:

First FMW Theorem

Flanagan, Marolf and Wald, [FMW00], reduce the first theorem to two conditions:

- Associated with each lightsheet in spacetime there is an entropy flux S_L^a whose integral over the lightsheet is the entropy flux through the lightsheet.
- The inequality $|S_L^a k^a| \leq \pi(\lambda_\infty - \lambda) T_{ab} k^a k^b$ holds everywhere on the lightsheet. Here λ_∞ is the value of the affine parameter at the endpoint of the lightsheet.

A consequence of demanding that the only modes that are fully captured on the lightsheet contribute to the entropy flux vector S_L^a results in it being defined non-locally. This gives the entropy a non-local character.

One would expect that if the lightsheet is only considered in part and not completely up to the caustic, resulting in a smaller area A' than the complete area A, the entropy would still satisfy the bound. This was proved by FMW from the first theorem and is called the strengthened form:

$$S \leq \frac{A - A'}{4}, \tag{3.34}$$

which is illustrated in Figure 3.14. This formulation does have some instances where it is either saturated or violated, [Bou02].

Second FMW Theorem

Flanagan, Marolf and Wald, [FMW00], use a complicated proof to show the second theorem is composed of the following two conditions:

- An absolute entropy flux vector field s^a can be used to approximate accurately the entropy content of spacetime.
- The inequalities

$$(s_a k^a)^2 \leq \frac{1}{16\pi} T_{ab} k^a k^b$$
$$|k^a k^b \nabla_a s_b| \leq \frac{1}{4} T_{ab} k^a k^b$$

  hold everywhere in spacetime, for any null vector k^a.

These conditions are satisfied by a wide range of matter systems. It is easy to test for these conditions e.g. in adiabatically evolving cosmologies. In these instances there is then no need for the whole procedure of constructing and testing lightsheets. Entropy flux assumes that the entropy is a local fluid which is a good approximation but ignores the non-local character.

3.12.3 Relation to other bounds

Is it possible to derive the spacelike entropy bound and Bekenstein bound from the covariant entropy bound? This would be possible if the covariant entropy bound is in fact a more general

and underlying formulation of these theories of limited applicability. To show that this is true is our objective for this section.

We begin by introducing the spacelike projection theorem by limiting the conditions of the covariant entropy bound. It can be stated in the following way: *Spacelike projection theorem* [Bou99]. *Consider a closed surface B that allows for at least one future directed lightsheet L. The lightsheet L is assumed to be* complete *where B is its only boundary. This is illustrated in Figure 3.15. Enclosed by B is the entropy $S(V)$ of the spatial region V on the same side as L. Then*

$$S(V) \leq S(L) \leq \frac{A}{4}. \tag{3.35}$$

Proof. All matter on V regardless of the choice of V or the time coordinates will pass through L. Thus the second law of thermodynamics is ensured by the first inequality while the last inequality is preserved by the covariant entropy bound.

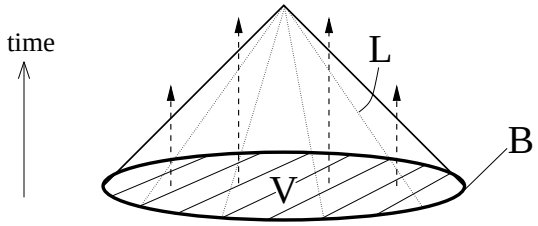


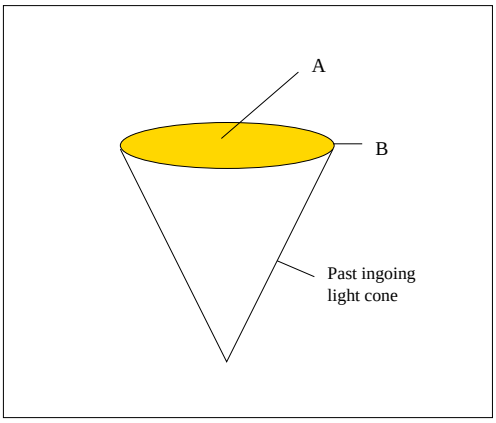
Figure 3.15: A surface, B, enclosing a spatial region, V, with a future directed lightsheet, L, implies the spacelike projection theorem. Taken from [Bou02].

For a closed weakly gravitating smooth surface we expect that the spacelike bound would be valid. It will only be in a strongly gravitating system that it would violate the spacelike bound. For example if there was no future directed lightsheet it would be anti-trapped and would imply that it was a strong gravitational system. Similarly if L had other boundaries, it would mean that there is a future singularity, which again means that we are dealing with strong gravity.

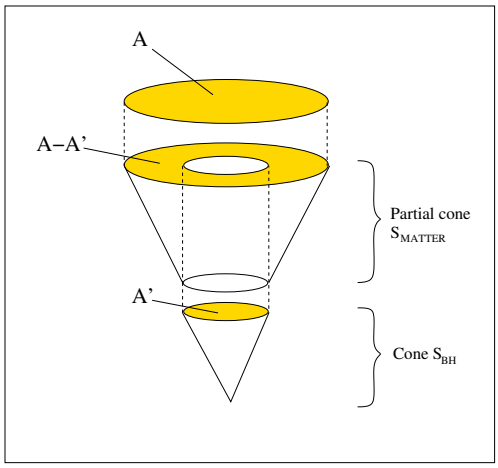
It turns out that if it wasn't for the spacelike projection theorem the use of the spherical entropy bound, which is needed for the generalised second law of thermodynamics in the Susskind process, Section 3.3, wouldn't be possible. The assumptions used in the discussion of the Susskind process are adhered to in the spacelike projection theorem.

Turning our focus now to the generalised second law of thermodynamics and the Bekenstein bound we first consider a black hole that is newly formed and bounded by B of area A coinciding with the event horizon. The past ingoing lightsheet contains all of the matter that formed the black hole. We can very simply conclude that $S_{\text{matter}} \leq A(B)/4 = S_{\text{BH}}$ by the covariant bound and this in turn satisfies the generalised second law. This is illustrated in the Figure 3.16(a).

In a more general setting where matter falls into a black hole we must consider previous discussions on matter falling into existing black holes, Section 3.2. If we consider the same scenario as previously described where a surface B is on the event horizon, it has an area of A'. When more matter falls into the black hole, the area increase from A' to A. The entropy added is S_{matter} and to an outside observer the Bekenstein entropy of the black hole has increased by $(A - A')/4$. The covariant bound does not by itself satisfy this process and the strengthened form, Eq. (3.34), of the covariant bound considered by FMW in Section 3.12.2, is required. The



(a) All the matter entropy, S_{matter}, contained on the past ingoing lightsheet contributes to the formation of the black hole and its eventual entropy, S_{BH}, where the event horizon coinciding with the bounding region B and area A. Therefore $S_{\text{matter}} \leq S_{\text{BH}} = A(B)/4$ and the covariant bound concurs with the generalised second law for black hole formation.



(b) When matter entropy, S_{matter} is added to the entropy of the existing black hole, $S_{\text{BHold}} = A'/4$, to form a new black hole with entropy, $S_{\text{BHnew}} = A/4$, the total entropy is thus $S_{\text{total}} = S_{\text{BHnew}} = S_{\text{BHold}} + S_{\text{matter}}$ where the light cone for the existing black hole has to be accounted for, shown at the bottom of the picture, along with the partial cone of the contributed matter entropy, shown in the middle of the diagram. The matter entropy is thus $S_{\text{matter}} = S_{\text{BHnew}} - S_{\text{BHold}} = A - A'/4$ which is the strengthened form of the covariant bound given by (3.34).

Figure 3.16: An illustration of black hole formation and matter being added to an existing black hole from the perspective of covariant entropy bound.

illustration in Figure 3.16(b) allows us to better understand the process described above.

The strengthened form of the covariant entropy bound can derive a form of the Bekenstein bound, but it is a very restrictive formulation and is violated by the use of Unruh radiation, [Unr76], in a process devised by Unruh and Wald, [UW82] and [UW83]. This is only applicable in very flat systems, [Bek83] and [Bek94]. The strengthened form is useful in certain instances but clearly isn't as generally applicable as the covariant entropy bound.

In conclusion the covariant entropy bound lays the foundation for the spherical bound and in turn the spacelike projects theorem under certain conditions. These allow for the preservation of the validity of the generalised second law for black hole formation. The strengthened form of the covariant entropy bound implies the generalised second law for absorption processes. It also yields the Bekenstein bound under very strict assumptions and illustrates that this form cannot be universally applicable.

3.13 Examples of applying the covariant entropy bound

In this section we test the covariant entropy bound in real spacetime settings and illustrate how it doesn't succumb to problems previously encountered in other entropy bound formulations. The first setting is cosmology where it is found to be well behaved under a variety of circumstances.

3.13.1 Cosmology

3.13.1.1 FRW metric and entropy density

We attempt to test the covariant entropy bound in a setting that is a good approximation to our own observable universe. For this we will use the Friedman-Robertson-Walker (FRW) metric, discussed in Section 2.12, remembering the metric

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2(dx^2 + dy^2 + dz^2)\right), \tag{3.36}$$

and using the conformal time, η, and comoving coordinates, χ,

$$d\eta = \frac{dt}{a(t)}, \quad d\chi = \frac{dr}{\sqrt{1-kr^2}}, \tag{3.37}$$

the FRW metric becomes

$$ds^2 = a^2(\eta)\left(-d\eta^2 + d\chi^2 + f^2(\chi)(dx^2 + dy^2 + dz^2)\right). \tag{3.38}$$

Open, flat and closed universe correspond to $k = -1, 0, 1$ and $f(\chi) = \sinh\chi, \chi, \sin\chi$ respectively, as mentioned previously.

The entropy in cosmological setting is described by a density per physical volume,

$$S(V) = \int_V d^3x\sqrt{h}\sigma. \tag{3.39}$$

The entropy density is only dependent on time and it decreases as the universe expands. We are assuming an adiabatic evolution of the universe,

$$\sigma(\eta) = \frac{s}{a(\eta)^3}, \tag{3.40}$$

where s is the comoving entropy density and remains constant at any space or time.

3.13.1.2 Expansion and apparent horizon

We now follow the discussion of Section 3.11.4 and try to determine the expansion of the orthogonal light rays, remembering that we are looking for at least two that will not expand ($\theta \leq 0$). We consider the expansion of the light rays orthogonal around a sphere and where we have taken the affine parameter to locally match $\pm 2\eta$, which is then used in Eq. (3.28). Adopting Bousso's notation the four light sheet directions F_1, F_2, F_3, F_4 are, respectively, represented by $(+,+)$, $(+,-)$, $(-,+)$ and $(-,-)$. The first sign refers to the time direction η and the second refers to larger or smaller spatial directions χ.

Using Eq. (3.28) and the above defined notation, we arrive at the following light sheet families,

$$\theta_{++} = \frac{\dot{a}}{a} + \frac{f'}{f}, \tag{3.41}$$

$$\theta_{+-} = \frac{\dot{a}}{a} - \frac{f'}{f}, \tag{3.42}$$

$$\theta_{-+} = -\frac{\dot{a}}{a} + \frac{f'}{f}, \tag{3.43}$$

$$\theta_{--} = -\frac{\dot{a}}{a} - \frac{f'}{f}. \tag{3.44}$$

The first term $(\dot{a}/a)$ in all of the above equations is positive if the universe is expanding and in turn negative if it is contracting. The second term is $\cot\chi$ $(1/\chi; \coth\chi)$ for a closed (flat; open) universe. There are different solutions for which the above equations would produce a $\theta \leq 0$ result. The apparent horizon is defined geometrically as a sphere at which at least one pair of orthogonal null congruence have zero expansion. The condition

$$\frac{\dot{a}}{a} = \pm\frac{f'}{f}, \tag{3.45}$$

identifies the location of the apparent horizon χ_{AH} as a function of time. There exists a single solution for the open and flat universes but there are generally two solutions which are symmetric about the equator.

We calculate the proper area of the apparent horizon by,

$$A_{\mathrm{AH}(\eta)} = 4\pi a(\eta)^2 f[\chi_{\mathrm{AH}}(\eta)]^2 = \frac{4\pi a^2}{\left(\frac{\dot{a}}{a}\right)^2 + k}. \tag{3.46}$$

Substituting the Friedman equation, discussed in Section 2.12,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\rho a^2}{3} - k, \tag{3.47}$$

we arrive at

$$A_{\mathrm{AH}}(\eta) = \frac{3}{2\rho(\eta)}, \tag{3.48}$$

where ρ is the energy density of matter.

- $A < A_{\mathrm{AH}}$ occurs when the term f'/f is dominant and are all normal surfaces. Since this term is dominant the cosmological evolution has no effect on the lightsheet directions. The two lightsheets will be for the past and future directed family going to the same spatial side. This corresponds to Eq. (3.42) and Eq. (3.44) where we can see how we arrived at these lightsheet families. In a flat or open universe, they will be directed towards $\chi = 0$ and in a closed universe it will be directed towards the nearest pole, $\chi = 0$ or $\chi = \pi$.
- $A > A_{\mathrm{AH}}$ occurs when the term $\dot{a}/a$ is dominant which results in two possibilities.
  - When $\dot{a} > 0$ we have an expanding universe (Big bang). From Eq. (3.43) and Eq. (3.44) we can conclude that both lightsheets would be past directed and are anti-trapped.

– When $\dot{a} < 0$ we have a collapsing universe (Big crunch). From Eq. (3.41) and Eq. (3.42) we can similarly conclude that both lightsheets are future directed families and will have negative expansion. This describes trapped spheres in a collapsing universe.

This is a complete classification for all the spherical surfaces in all FRW universes according to their lightsheet directions.

3.13.1.3 Lightsheet vs. spatial volumes

Before calculating the entropy on the lightsheets we review again the failure of the spacelike entropy bound identified in Section 3.10. The area of a sphere at η_0, χ_0 is given by

$$A(\eta_0, \chi_0) = 4\pi a(\eta_0)^2 f(\chi_0)^2. \tag{3.49}$$

Comparing this area to the enclosed spatial volume $V(\chi_0)$ of entropy, where $\chi \leq \chi_0$ at equal time $\eta = \eta_0$ and remembering that we have assumed an adiabatic evolution of the universe we only require χ_0, since the entropy remains constant for χ_0,

$$S[V(\chi_0)] = 4\pi s \int_0^{\chi_0} d\chi f(\chi)^2. \tag{3.50}$$

In a flat universe $[f(\chi)] = \chi$, the area grows as χ_0^2 while the entropy grows like χ_0^3. As pointed out earlier, Section 3.10.2, this results in a violation where $S[V(\chi_0)] > A$ for sufficiently large χ_0. In a closed universe, where $(f(\chi) = \sin\chi)$, χ ranges from 0 to π. As $A \to 0$ for $\chi_0 \to \pi$ this leads to the violation covered in Section 3.10.1. In general the entropy is increasing monotonically in this range. Lightsheets overcome this problem by always being mindful of the contraction condition, $\theta \leq 0$, which ensures that the entropy area ratio doesn't diverge for large χ but instead levels out at a constant value.

3.13.1.4 Solution with fixed equation of state

As discussed in Section 3.12.2 the FRW universe can be described by a perfect fluid, with stress tensor

$$T^a_{\ b} = \mathrm{diag}(-\rho, \rho, \rho, \rho). \tag{3.51}$$

A fixed equation of state is assumed,

$$p = w\rho. \tag{3.52}$$

This allows for a very general solution in that it is able to generate many alternate scenarios by allowing a rapid transition between different effective equations of state. We can better understand this statement if we look at the history of our universe where for most of its lifetime it was dominated by a pressureless dust, $w = 0$. Radiation was dominant during the early stages of the universes life cycle, $w = \frac{1}{3}$. A cosmological constant corresponds to $w = -1$ and should be taken into account for the very early times and possibly the present.

We can determine the scale factor for Eq. (3.38) by using the above assumptions and condi-

tions,

$$a(\eta) = a_0 \left[f\left(\frac{\eta}{q}\right) \right]^q, \tag{3.53}$$

where

$$q = \frac{2}{1+3w} \tag{3.54}$$

and f is equal to either sin, I (the identity) or sinh for closed, flat or open universes respectively, as in Eq. (3.38). It follows from Eq. (3.45) that an apparent horizon is located at

$$\chi_{\mathrm{AH}}(\eta) = \frac{\eta}{q}, \tag{3.55}$$

for all instances. There is also a mirror horizon at $\pi - \frac{\eta}{q}$ in the closed space case.

3.13.1.5 Flat universe

We now explore all possible lightsheets of all spherical areas ($0 < \chi < \infty$) at a time η in a flat FRW universe,

$$A(\eta, \chi) = 4\pi a(\eta)^2 \chi^2. \tag{3.56}$$

As highlighted in the evaluation of the lightsheets $\chi \leq \chi_{\mathrm{AH}}(\eta)$ means the sphere is normal and the lightsheet direction is $(+-)$ and $(--)$, while for $\chi \geq \chi_{\mathrm{AH}}$ the sphere is anti-trapped with lightsheets $(-+)$ and $(--)$. Looking at the future ingoing lightsheet first, it contracts towards the origin and generates a conical lightsheet whose coordinates (χ', η') obey

$$\chi' + \eta' = \chi + \eta. \tag{3.57}$$

The comoving entropy in the region $0 \leq \chi' \leq \chi$ on this lightsheet is given by

$$S_{+-} = \frac{4\pi}{3} s \chi^3. \tag{3.58}$$

If we consider the ratio of entropy to the area we arrive at,

$$\frac{S_{+-}}{A} = \frac{s\chi}{3a(\eta)^2}. \tag{3.59}$$

The above ratio can be maximised by the outermost normal surface at any given time η. This is the sphere on the apparent horizon. Thus we obtain the bound

$$\frac{S_{+-}}{A} \leq \frac{s\chi_{\mathrm{AH}}(\eta)}{3a(\eta)^2}. \tag{3.60}$$

Next considering the past-ingoing $(--)$ lightsheet of any surface with $\chi < \eta$ also terminates at a caustic $\chi = 0$. The lightsheet is truncated instead by the big bang singularity at $\eta = 0$ if $\chi > \eta$. In this instance it will contain the comoving entropy in the region $\chi - \eta \leq \chi' \leq \chi$. Looking at the the entropy to area ratio,

$$\frac{S_{--}}{A} = \frac{s\eta}{a(\eta)^2} \left(1 - \frac{\eta}{\chi} + \frac{\eta^2}{3\chi^2}\right). \tag{3.61}$$

By maximising the above ratio for large spheres, ($\chi \to \infty$), we arrive at a bound

$$\frac{S_{--}}{A} \leq \frac{s\eta}{a(\eta)^2} \tag{3.62}$$

for the $(--)$ lightsheets at time η.

Lastly we turn our attention to the past-outgoing $(-+)$ lightsheet of any surface with $\chi > \chi_{\rm AH}$. It is truncated by the big bang singularity and contains the entropy within $\chi \leq \chi' \leq \chi + \eta$. The ratio of the entropy to the area,

$$\frac{S_{-+}}{A} = \frac{s\eta}{a(\eta)^2}\left(1 + \frac{\eta}{\chi} + \frac{\eta^2}{3\chi^2}\right) \tag{3.63}$$

is maximised for the smallest possible value of χ which is the apparent horizon. The bound can be determined as being

$$\frac{S_{-+}}{A} \leq \frac{s\eta}{a(\eta)^2}\left(1 - \frac{\eta}{\chi_{\rm AH}} + \frac{\eta^2}{3\chi_{\rm AH}^2}\right). \tag{3.64}$$

Using the solution for the fixed equation of state and setting $a_0 = 1$ for convenience,

$$a(\eta) = \left(\frac{\eta}{q}\right)^q, \qquad \chi_{\rm AH}(\eta) = \frac{\eta}{q}. \tag{3.65}$$

We finally arrive at the following answer where the bound for all three types of lightsheets at anytime are given by

$$\frac{S}{A} \leq s\eta^{1-2q}, \tag{3.66}$$

up to factors of order unity.

Assuming that a Planck volume contains no more than one bit of information and being able to deduce that at a single Planck volume the entropy is roughly given by s, from Eq. (3.40) as a first order approximation, we can conclude that $s \sim 1$ at the Planck scale.

This enforces the covariant bound given in Eq. (3.31) and if we apply the above reasoning to Eq. (3.66). This bound is persevered for all lightsheets at later times $\eta > 1$ if $q \geq \frac{1}{2}$. This translates to the condition that $w \leq 1$ from Eq. (3.54), this was refined in [KL99] for a more general setting to be $-1 < w \leq 1$. Thus for any flat FRW universe the covariant entropy bound is satisfied as long as the equation of state satisfies $-1 < w \leq 1$. This is a very reasonable physical state which prohibits superluminal flow of energy.

3.13.1.6 A cosmological corollary

What is the largest volume in a cosmological spacetime that the spacelike holographic principle can be applied to? The spacelike projection theorem, Section 3.12.3, guarantees that the spacelike entropy bound will hold for surfaces that admit a future directed complete lightsheet. In cosmology surfaces on or within the apparent horizon are normal and hence have a future directed lightsheet, but aren't necessarily complete. We thus arrive at the following corollary to the spacelike projection theorem, [Bou99]: *The area of any sphere within the apparent horizon exceeds the entropy enclosed in it, if the future lightsheet of the sphere is complete.*

3.13.2 Gravitational Collapse

Inside a collapsing regime one can no longer rely on the generalised second law. It has already been shown that in the case of a collapsing star, Section 3.10.3, and collapsing universe, Section 3.10.2, we encounter serious problems in trying to resolve these instances.

The simplest way to show that the covariant entropy bound is valid, is to substantiate the local hypotheses of Flanagan, Marolf and Wald, [FMW00]. If this isn't possible then one must explicitly find the lightsheets. We will examine the bound in two examples, a collapsing star and a collapsing universe, remembering that the general entropy bounds, like the spacelike entropy bound, do not hold in these cases.

3.13.2.1 Collapsing universe

A good simple starting point is to consider the adiabatic recollapse of a closed FRW universe. This is the reverse of what we looked at in the previous Section of an expanding FRW universe and it is easy to see that we will once again satisfy the covariant bound in the same way since this is a unitary process and time reversal is possible. This would simply change the direction of the lightsheets discussed previously. Small spheres near the poles are normal while larger spheres which are anti-trapped during expansion will now become trapped during collapse. Their lightsheets are future directed and truncated by the future singularity (Big crunch) in order to preserve the bound.

The calculation for anti-trapped spheres was done by Fischler and Susskind, [FS98], which follows a similar procedure as that of the flat case, Section 3.13.1.5. Remembering the complications of small spheres raised in Section 3.13.1.3 it is wise to be mindful of this during the analysis.

3.13.2.2 Collapsing star

The Oppenheimer-Snyder solution, [MTW73] and [HE73], describes very accurately the metric in and around a collapsing star. This solution uses a suitable portion of the collapsing closed FRW universe to model the star. In order to achieve this we focus in on the coordinate range

$$0 \leq \chi \leq \chi_0, \eta > q\frac{\pi}{2}, \tag{3.67}$$

for the metric of Eq. (3.38) and note that $\chi_0 < \pi/2$. The outside of the star is empty and by Birkhoff's theorem the metric will be given by a portion of the Schwarzschild solution Eq. (2.64).

The lightsheet directions can be deduced from the corresponding closed universe, as used in the FRW case, and for the Schwarzschild solution. This is illustrated in the Penrose diagram 3.17. When the apparent horizon coincides with the surface of the star at sufficiently late times, a black hole forms. This means that the star's surface is forever trapped at later times and only results in future directed lightsheets near the future singularity. From Eq. (3.49)

$$A(\chi_0, \eta) = A_{\max}\left(\sin\frac{\eta}{q}\right)^q, \tag{3.68}$$

is the area of a star. The future singularity is at the time $\eta = q\pi$ while the area at maximum

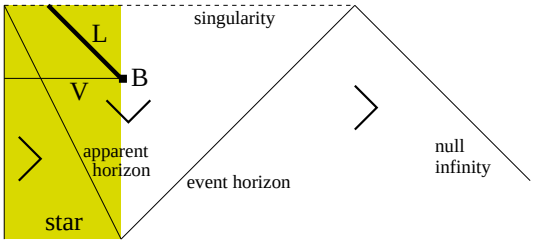


Figure 3.17: Penrose diagram of collapsing star. The shaded region is the spatial region of the star. L is a Lightsheet for the region bounded by B. V is a spatial slice for the bounded region B. Taken from [Bou02].

expansion is $A_{\max} \equiv 4\pi a_0^2 \sin^2 \chi_0$.

Recalling the discussion of the spacelike bounds violation for a collapsing star, Section 3.10.3, we are reminded that, at very late times $\eta_0 \to \pi$, the surface area approaches zero, $A(\chi_0, \eta_0) \to 0$ as we approach the singularity. This results in the violation of the spacelike entropy bound $S(V) > A(B)$ since the entropy of the star doesn't decrease. The entropy $S(L)$ of the lightsheet, L, does however manage to overcome a similar violation by vanishing in this limit. This occurs since the future singularity truncates the lightsheet and thus results in it only passing through a limited portion of the star and thus only represents a corresponding portion of the star's entropy. The lightsheet can only capture a representative portion of entropy to which it has been "exposed" for a given bounded area B. In this instance the singularity cuts short its full "exposure" to the stars entropy.

This can be seen in Figure 3.17 where at late times the area becomes very small for B. The lightsheet, L, only crosses a portion of the shaded region, which is the spatial region of the star, because it is shortened by the singularity. The spatial region V remains the same at all times and the corresponding entropy thus remains finite resulting in a violation of the spacelike entropy bound as we approach the singularity.

3.13.2.3 Collapsing shell

Take a small black hole of radius $r_0 = 2m$ with an apparent horizon that is a null hypersurface with spacelike spherical cross section of area $A = 4\pi r_0^2$. Choose a sphere B of area A on the apparent horizon, then by definition, the expansion of the past directed ingoing and the future directed outgoing light rays vanish near B, so both are allowed lightsheet directions. The past directed ingoing lightsheet contains all the entropy that formed the black hole S_{orig}. This implies that the covariant bound for the past case reduces to the generalised second law where the horizon entropy, $A/4$, is greater than the matter entropy that formed the black hole, S_{orig}.

We are interested in the future directed lightsheet. This lightsheet will continue forever at zero expansion along the apparent horizon as long as no more matter falls into the black hole. If more matter does fall into the black hole it will cause the horizon to grow $r > r_0$, but this will result in the lightsheet L collapsing due to Eq. (3.32). Again the fact that the lightsheet eventually reaches the singularity ensures that it only encounters a finite amount of entropy less than $A/4$.

We can construct a symmetric shell surrounding the black hole with a mass m and width

w. As long as the shell's radius is very large in comparison with that of the black hole's, the local gravitational effects can be suppressed in order for the shell's mass and width to be chosen arbitrary. With a symmetric and weakly gravitating system, put in place by the above condition, the Bekenstein bound, $S_{\text{matter}} \leq 2\pi ER$ where E is the total mass energy of a matter system, is well suited to this scenario and has good empirical evidence supporting it, [Bek81] and [Bek84].As a consequence of excitations being carried by radial modes it is possible to divide the shell along radial walls. Several weakly gravitating systems result but none have a larger length scale than w. The Bekenstein's bound applies to each of these and after reassembling the shell the total entropy is bounded by

$$S \leq 2\pi Mw. \tag{3.69}$$

There is no restriction on the value of M or w, so the entropy contained in this system is arbitrarily large.

Now we consider an adiabatic collapse of this shell onto the black hole. At some point the inner portion of the shell will come into contact with the lightsheet being generated by L. Let the area of the shell at this point be A. Once the light rays pass through the shell walls and are exposed to the matter of the shell they begin to focus and form a caustic. To ensure that the caustic only forms after it has penetrated the entire shell's surface, the shell's mass and width are restricted to

$$Mw \leq r_0^2/2. \tag{3.70}$$

Now applying this restriction to Eq. (3.69) produces

$$S \leq \pi r_0^2 = \frac{A}{4}. \tag{3.71}$$

We can thus see that the lightsheet L may saturate the covariant bound, but it is not violated.

3.14 Holographic Principle?

Although the covariant entropy bound is not the complete formulation to the Holographic Principle, it does have certain characteristics that have been developed to support it and will surely become the building blocks to ultimately produce a Holographic Principle. These are reviewed below:

- The covariant entropy bound and lightsheets are well defined. There are clear constructs for lightsheets and calculating the entropy related to an area.
- It is widely tested and applicable and, as yet, no physically realistic counter has been found.
- The result of the covariant entropy bound are very insightful and prove to be counter intuitive relating the area and not the volume to entropy.
- The covariant bound is not explained by other laws and is also not a consequence of black hole thermodynamics. This points to it being an underlying principle in physics or at least a clue to something more encompassing.

- Statistical entropy is used by the bound and there is no reliance on microscopic states. This gives a unique tool for uncovering the interplay between a classical concept and that of a quantum theory.
- This bound is not a consequence of black holes, but is, however, closely related to these exotic creations. The covariant bound can, however, be used to construct other bounds which are essential to the generalised second law which relates directly to black hole thermodynamics.
- The over all simplicity and generality of this principle is compelling and warrant more investigation.

Conclusion

To review what has been covered in this thesis, we began with a basic mathematical framework which introduced spacetime, changing coordinates, tensors and varying basis. This gave a platform upon which to introduce general relativity in the next chapter along with some more mathematical concepts. This chapter concluded with Einstein's field equations and some common solutions for these equations. The Newtonian theory of gravity was used to contrast with the theory of general relativity throughout the progression of the second chapter. The concluding chapter begins by introducing an example by 't Hooft which shows the link between the area of a black hole and its area. This then lays the president for exploring the birth and progression of entropy bounds relating the area and entropy of a bounded region. This highlights the conjecture of the Holographic Principle and examines possible working scaled down theories of such a form. The covariant entropy bound is then scrutinised in light of it being the most successful and general entropy bound.

The most significant implication found thus far in our discussion is that there exists an undeniable correspondence between geometric considerations of a region of spacetime, area, and the associated number of degrees of freedom to the respective region. This is a unexpected result and shifts the paradigm of how we think of information. This relationship is closely tied to null hypersurfaces or lightsheets and was conceived in the setting of black hole thermodynamics.

Let us consider what the implication are for such a relationship, as described above, has in terms of quantum field theory. It is clear that the one thing the Holographic Principle points to is that we are dealing with a situation where we have a smaller than anticipated, in some cases even a finite dimensional, Hilbert space. In these cases this would imply that we can no longer accept the canonical commutation relations in the same light as we have previously done since this always assumed an infinite dimensional Hilbert space. At least in these cases we thus have to develop a new means of interpreting the canonical commutation relations or develop a new quantisation prescript. One possibility that takes account of this may be some of the non-commuting quantum field theory's being explored at the moment.

There are two lines of attack for developing the Holographic Principle. Physics has very strong local characteristics and one can thus accordingly try and cast a holographic theory from this perspective, [tH99], [tH00], [tH01a], [tH01b] and [tH01c]. The challenge here is to develop a gauge invariance retaining the physical degrees of freedom. This approach has as yet not shown an emergence of the area relation relative to the number of degrees of freedom, [Bou02]. The second approach takes this fact, the area relation to the number of degrees of freedom, as the departure point and regards locality as an emergent phenomenon. Bousso follows this approach in the body of his work. The main stumbling block here is to try and understand how locality does arise.

The biggest criticism of the covariant entropy bound is that it has no fundamental basis or underlying law rooted in our existing understanding of physics, [FMW00]. This doesn't necessarily discredit its validity but rather makes it difficult to understand and justify its success. An important step in solidifying the place of the covariant entropy bound, as a steering principle

in quantum gravity, is if more experimental data can be found to support it. This could be found in a cosmological setting. Another avenue is if the covariant entropy bound can be stated in a different form where it can be shown that it is a consequence of underlying conditions of spacetime and physics. This would be similar to the idea introduced by Flanagan, Marlof and Wald, [FMW00], where conditions ensure that the covariant entropy bound works but would be justified from a more fundamental stand point.

There have also been very positive results that have lent support for the holographic principle. These range from the semi-classical calculation of Hawking radiation where the result $A/4$ can be exactly determined and also the famous AdS/CFT correspondence where the number of degrees of freedom agree manifestly with the Holographic Principle, [Mal98]. There is limited success in trying to draw string theory and Holography together. There are several attempts in string theory to examine the relationship of non-locality exhibited by the Holographic Principle, two examples of which are [LSU94] and [LP$^+$94]. A perturbative string theory does however appear to suffer from manifest failings in agreeing with Holography, [Sus95], where the perturbative expansion breaks down while testing the violation of the bound. If Holography could be put on a more sound footing it would be useful as a guiding principle to further develop and bring together a more comprehensive structure for string theory.

The Holographic Principle may have been an audacious concept to propose but in the light of intervening developments has become a difficult one to ignore and requires more investigation. Upon reflection it may be that Plato's words will echo true in the setting of modern physics:

"*To them, I said, the truth would be literally nothing but the shadows of the images*".

Bibliography

[Bek72] J.D. Bekenstein. Black holes and the second law. *Nuovo Cim. Lett.*, (4):737, 1972.

[Bek73] J.D. Bekenstein. Black holes and entropy. *Phys. Rev.*, D(7):2333, 1973.

[Bek74] J.D. Bekenstein. Generalized second law of thermodynamics in black hole physics. *Phys. Rev.*, D(9):3292, 1974.

[Bek81] J.D. Bekenstein. A universal upper bound on the entropy to energy ratio for bounded systems. *Phys. Rev.*, D(23):287, 1981.

[Bek83] J.D. Bekenstein. Entropy bounds and the second law for black holes. *Phys. Rev.*, D(27):2262, 1983.

[Bek84] J.D. Bekenstein. Entropy content and the second law for black holes. *Phys. Rev.*, D(27):2262, 1984.

[Bek94] J.D. Bekenstein. Do we understand black hole entropy? *gr-qc/9409015*, 1994.

[Bou99] R. Bousso. A covariant entropy conjecture. *JHEP*, 06(004), 1999.

[Bou00] R. Bousso. The holographic principle for general back grounds. *Class. Quant. Grav.*, 17:997, 2000.

[Bou02] R. Bousso. The holographic principle. *Rev. Mod. Phys.*, 74:825–874, 2002.

[Car23] É. Cartan. Sur les variétés à connexion et la théorie de la relativité generalisée (premiére partie). *Ann. École Norm. Sup.*, (40):325–412, 1923.

[Car24] É. Cartan. Sur les variétés à connexion et la théorie de la relativité generalisée (suite). *Ann. École Norm. Sup.*, (41):1–25, 1924.

[Ell99] G.F.R. Ellis. Course on general relativity. Lecture notes in Mathematics and Applied Mathematics, U.C.T., 1999.

[FMW00] E.E. Flanagan, D. Marolf, and R.M. Wald. Proof of classical version of Bousso entropy bound and the generalized second law. *Phys. Rev.*, D(56):4922, 2000.

[FS98] W. Fischler and L. Susskind. Holography and cosmology. *hep-th/9806039*, 1998.

[Har03] J.B. Hartle. *Gravity: An Introduction to Einstein's General Relativity.* Addison Wesley, San Francisco, 2003.

[Haw71] S. W. Hawking. Gravitational radiation from colliding black holes. *Phys. Rev. Lett.*, (25):152, 1971.

[Haw72] S. W. Hawking. Black holes in general relativity. *Commun. Math. Phys.*, (43):152, 1972.

[Haw74] S. W. Hawking. Black hole explosions. *Nature*, (248):30, 1974.

[Haw75] S. W. Hawking. Particle creation by black holes. *Commun. Math. Phys.*, (43):199, 1975.

[HE73] S.W. Hawking and G.F.R. Ellis. *The Large Scale Structure of Space-Time.* Cambridge University Press, Cambridge, 1973.

[KL99] N. Kaloper and A. Linde. Cosmology vs. holography. *hep-th/9904120*, 1999.

[Low99] D. A. Lowe. Comments on a covariant entropy bound conjecture. *JHEP*, (10):026, 1999.

[LP+94] D.A. Lowe, , J. Polchinski, L. Susskind, L Thorlacius, and J. Uglum. Black hole complementarity versus locality. *Phys. Rev.*, D(52):6997, 1994.

[LSU94] D.A. Lowe, L. Susskind, and J. Uglum. Information spreading in interacting string field theory. *Phys. Lett.*, B327(226), 1994.

[Mal98] J. Maldacena. The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.*, (2):231, 1998.

[MTW73] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation.* W.H. Freeman and Company, San Francisco, 1973.

[Smo01] L. Smolin. The strong and weak holographic principles. *Nucl. Phys.*, B601:209, 2001.

[Sus95] L. Susskind. The world as a hologram. *J.Math.Phys.*, (36):6377, 1995.

[TE99] R.K. Tavakol and G.F.R. Ellis. On holography and cosmology. *Phys. Lett.*, B469, 1999.

[tH85] G. 't Hooft. On the quantum structure of a black hole. *Nucl. Phys.*, B(256):727, 1985.

[tH93] G. 't Hooft. Dimensional reduction in quantum gravity. *hep-th/9310026*, 1993.

[tH99] G. 't Hooft. Quantum gravity as a dissipative deterministic system. *gr-qc/9903084*, 1999.

[tH00] G. 't Hooft. Determinism and dissipation in quantum gravity. *hep-th/0003005*, 2000.

[tH01a] G. 't Hooft. Determinism in free bosons. *hep-th/0104080*, 2001.

[tH01b] G. 't Hooft. How does god play dice? (Pre-) determinism at the Planck scale. *hep-th/0104219*, 2001.

[tH01c] G. 't Hooft. Quantum mechanics and determinism. *hep-th/0104219*, 2001.

[Unr76] W.G. Unruh. Notes on black hole evaporation. *Phys. Rev.*, D(14):870, 1976.

[UW82] W.G. Unruh and R.M. Wald. Acceleration radiation and the generalised second law of thermodynamics. *Phys. Rev.*, D(25):942, 1982.

[UW83] W.G. Unruh and R.M. Wald. Entropy bounds, acceleration radiation, and the generalised second law. *Phys. Rev.*, D(27):2271, 1983.

[Wal84] R.M. Wald. *General Relativity.* The University of Chicago Press, Chigaco, 1984.